Delivering Data Reduction Pipelines to Science Users

Wolfram Freudling
Science Data Products Group
European Southern Observatory

Science Data Reduction and Operations

- A science data reduction system (DRS) is necessary for operations:
  - Quality control
  - Investigation of data quality and instrument performance
  - Trouble shouting
  - Production of archive products
- Scientists also need a DRS:
  - Customized data reduction
  - Improve on archival data
- Different requirements for different use cases
  - Unsupervised, robust “pipelines”
  - Interactive data reduction

ESOReflex: Requirements for interactive data reduction system

- Easy, simple, intuitive, interactive user interface
- Run pre-existing algorithms recipes in defined sequence
- Automatic organization of data
- Automatic bookkeeping
- feed them with necessary data
- Possibility to include user-supplied applications
- Visualize results
- Interaction with recipes
- Record and explore provenance
- Possibility to modify and share reduction sequence

Scientific Workflow Systems

- System to express multi-step computational tasks
- Workflows describe dependencies between tasks
- Several workflow engines on the market, including free NSF funded public-domain software.
- ESOReflex uses Kepler
  - Kepler is freely available under the BSD License.
  - https://kepler-project.org
  - Used in life science, ecology, geology
  - Kepler provides a graphical user interface (Java)
  - run--time engine that can execute workflows either from within the graphical interface or from command line
  - Current Kepler version 2.5

File Edit View Workflow Tools Window Help

Components | Data | Outline

Search Components

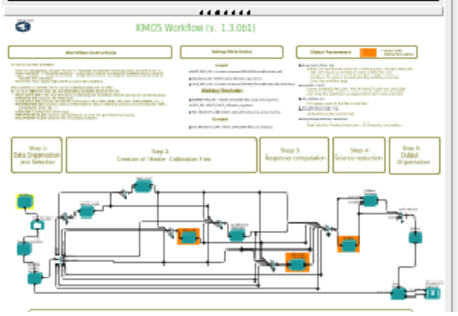
Search

Advanced ... | Sources | Cancel

All Ontologies and Folders

- Components
- Projects
- Statistics
- Demos
- Actors
- Dataturbine
- Directors
- Eso-reflex
- Job
- Opendap
- Outreach
- R

0 results found.



Workflow

DDF Director

KMOS Workflow (v. 1.3.0b1)

Workflow Instructions

In orer to run this workflow:

- Turn on highlighting. Choose "Tools"-> "Animate at Runtime" from top menu and set it to "1".
- Open "Window" -> "Runtime Window" in top menu before starting the workflow if you wish to monitor the reduction.
- Press the "Run" button OR cntrl-R to start the workflow.

The workflow is initially set to run on a default data set. In order to run on a different data set, the following variables have to be set:

- ROOT_DATA_DIR is the root directory containing the workflow related directories defined below
- RAWDATA_DIR contains the RAW data.
- CALIB_DATA_DIR contains the STATIC calibration files (REF_LINES, ARC_LIST, WAVE_BAND, etc...)
- BOOKKEEPING_DIR contains various informations about reduction process (esorex.cfg file, SOFs, parameters used, etc...)
- LOGS_DIR contains the esorex logs.
- TMP_PRODUCTS_DIR contains the products as they are generated by esorex.
- END_PRODUCTS_DIR contains the renamed products.

Setup Directories

Input:

- ROOT_DATA_DIR: /scratch/mneeser/SDP/KMOS/install/bin/data_wkf/
- RAW_DATA_DIR: $ROOT_DATA_DIR/reflex_input/kmos
- CALIB_DATA_DIR: /scratch/mneeser/SDP/KMOS/install/bin/install/calib/kmos-...

Working Directories:

- BOOKKEEPING_DIR: $ROOT_DATA_DIR/reflex_book_keeping/kmos
- LOGS_DIR: $ROOT_DATA_DIR/reflex_logs/kmos
- TMP_PRODUCTS_DIR: $ROOT_DATA_DIR/reflex_tmp_products/kmos

Output:

- END_PRODUCTS_DIR: $ROOT_DATA_DIR/reflex_end_products

Global Parameters = actor with interactive option

- RecipeFailureMode: Ask
  Defines the workflow behaviour for a failing recipe. Possible values are:
  - Ask: the choice to continue or stop is left to the user
  - Continue: the error is ignored and the workflow continues
  - Stop: the workflow stops
- EraseDirs: FALSE
  Erases BOOKKEEPING_DIR, TMP_PRODUCTS_DIR and LOGS_DIR each time the workflow is run (Lazy Mode won't work anymore)
- FITS_VIEWER: qfv
  FITS viewer used for the files inspection
- GlobalPlotInteractivity: True
  Global Interactivity control Point
- SelectDatasetMethod: Interactive
  Data Selection Method (Interactive, All, Complete, Incomplete)

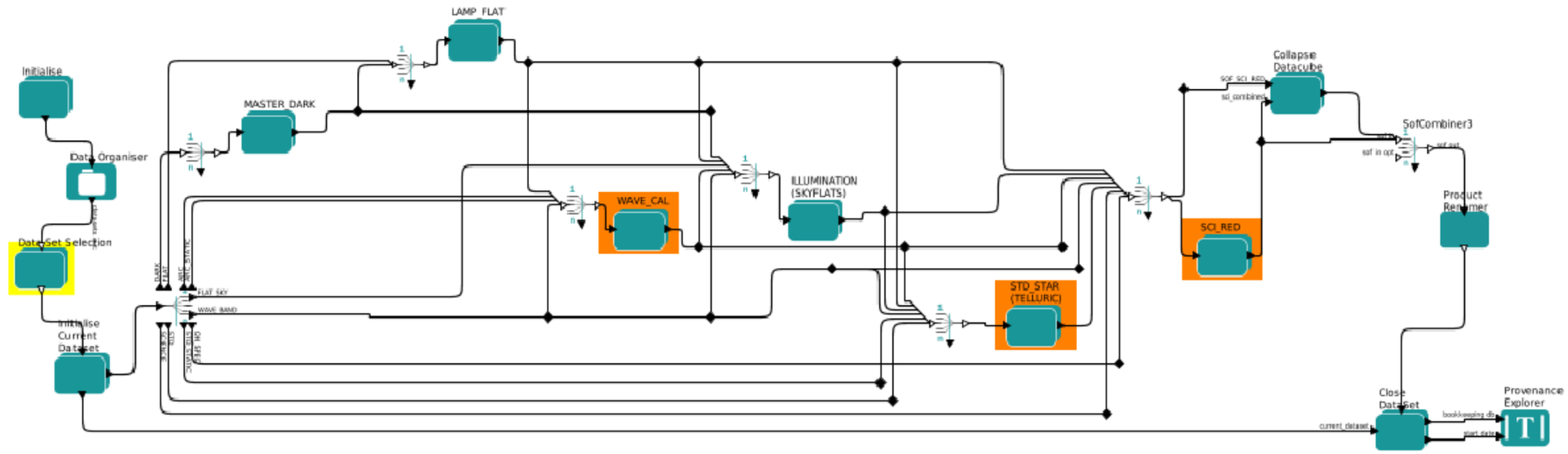
Step 1: Data Organisation and Selection

Step 2: Creation of Master Calibration Files

Step 3: Response computation

Step 4: Science reduction

Step 5: Output Organisation



Auxiliary and debug parameters, please do not change: GLOBAL_TIMESTAMP: 2014-03-05T16:02:45 | ESORexArgs: --suppress-prefix=TRUE | END_PRODUCTS_SUBDIR: 2014-03-05T13:40:20/KMOS.2013-06-30T23:48:06.049_tpl | N_SELECTED_DATASETS: 1

Workflow driven Data Reduction

A&A 559, A96 (2013)

A&A 559, A96 (2013)
DOI: 10.1051/0004-6361/201322494

Astronomy & Astrophysics

Automated data reduction workflows for astronomy

The ESO Reflex environment

W. Freudling, M. Romaniello, D. M. Bramich, P. Ballester, V. Forchi, C. E. García-Dabló, S. Moehler, and M. J. Neeser

European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
e-mail: wfreudli@eso.org

Received 6 August 2013 / Accepted 14 October 2013

ABSTRACT

Context. Data from complex modern astronomical instruments often consist of a large number of different science and calibration files, and their reduction requires a variety of software tools. The execution chain of the tools represents a complex workflow that needs to be tuned and supervised, often by individual researchers that are not necessarily expert for any specific instrument.
Aims. The efficiency of data reduction can be improved by using automatic workflows to organise and execute a sequence of data reduction steps. To realize such efficiency gains, we designed a system that allows intuitive representation, execution and modification of the data reduction workflow, and has facilities for inspection and interaction with the data.
Methods. The European Southern Observatory (ESO) has developed Reflex, an environment to automate data reduction workflows. Reflex is implemented as a package of customized components for the Kepler workflow engine. Kepler provides the graphical user interface to create an executable flowchart-like representation of the data reduction process. Key features of Reflex are a novel concept for a data organiser, infrastructure to re-use results, thorough book-keeping, data progeny tracking, interactive user interfaces, and automatic support concept to exploit information created during data organisation for the workflow execution.
Results. Automated workflows can greatly increase the efficiency of astronomical data reduction. In Reflex, workflows can be run non-interactively as a first step. Subsequent optimization can then be carried out while transparently re-using all unchanged intermediate products. We found that such workflows enable the reduction of complex data by non-expert users and minimizes mistakes due to book-keeping errors.
Conclusions. Reflex includes novel concepts to increase the efficiency of astronomical data processing. While Reflex is a specific

Reflex tutorials
http://eso.org/reflex

Pipeline manuals

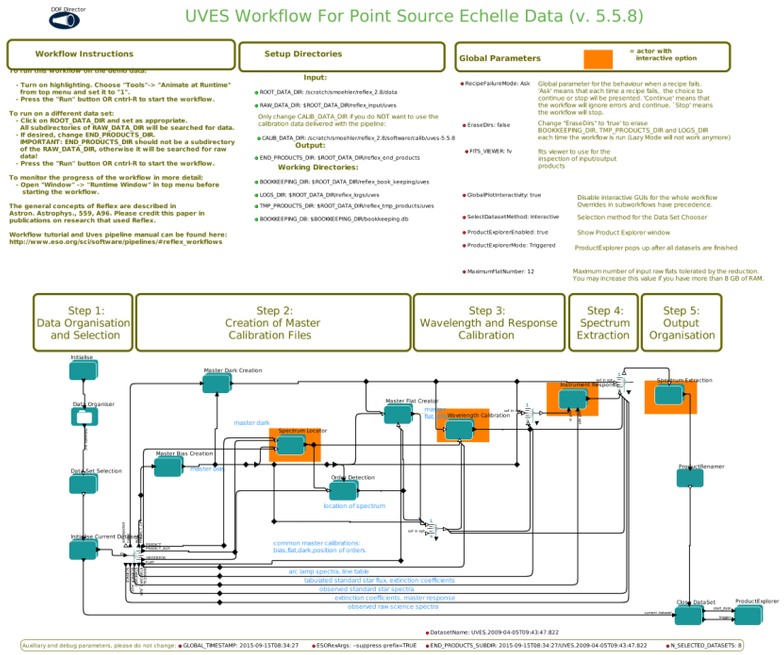
Forchi, V., Reflex User Manual, VLT-MAN-ESOs-19000-5037

ESOReflex Implementation

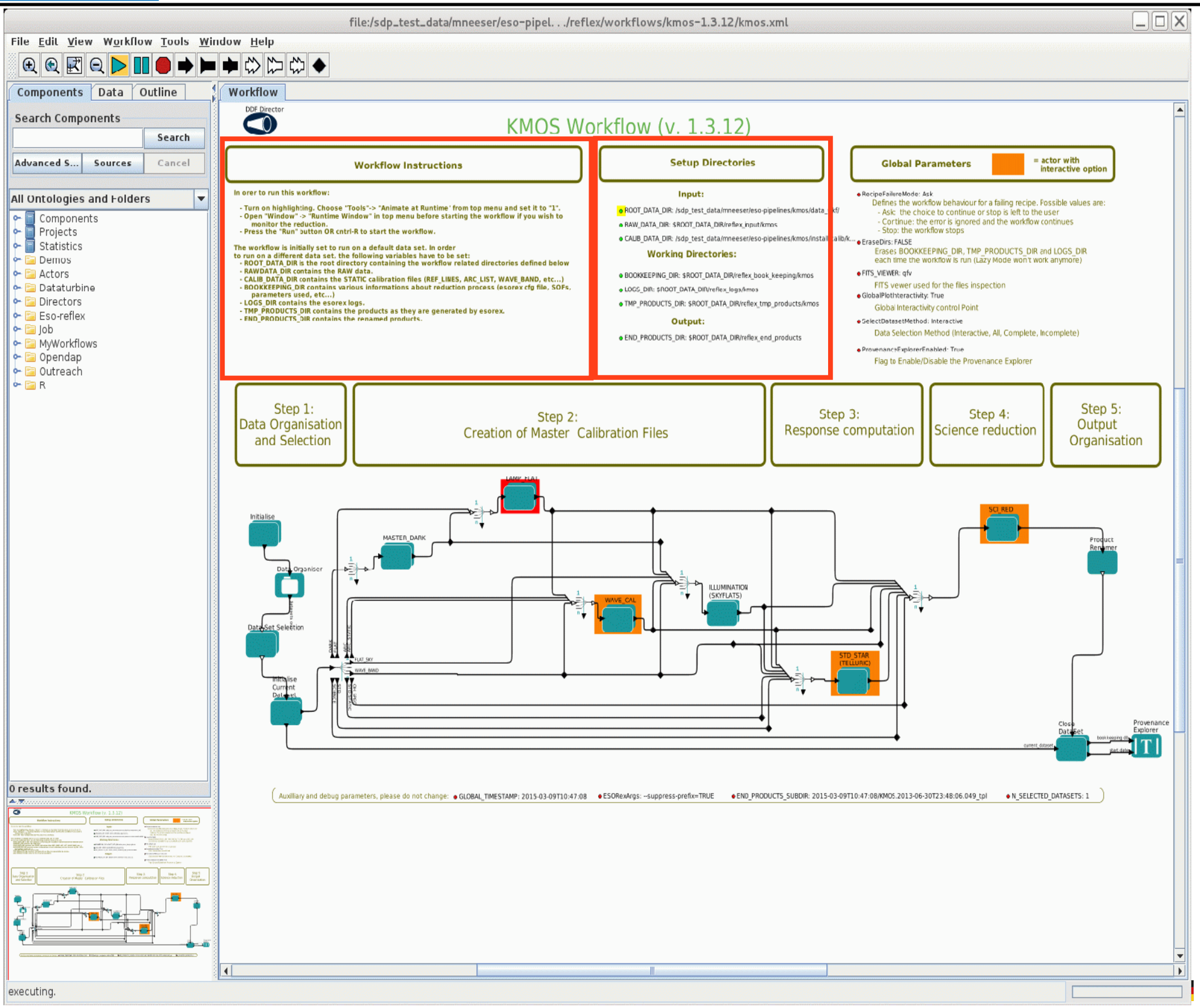
- Recommended environment to run ESO VLT pipeline “recipes” for users external to ESO
- ESO Reflex Team:
  - Project Scientist: Wolfram Freudling
  - Implementation of Reflex Environment: Vincenzo Forchi
  - Workflows implemented @ ESO:
    - Specification by Instrument teams Science Data Products group
    - Implementation by Pipeline Software Systems, Project Manager: Enrique Garcia
- Workflows for all new instruments required from instrument consortia

ESOReflex Key features

- Intuitive workflow design that shows the top level recipes and their interdependence
- Built-in rule based data organization
- Efficiency and speed:
  - Little overheads on top of running the recipes
  - Get “first science results” as soon as possible
  - Skip unnecessary steps
- Fully supported batch mode

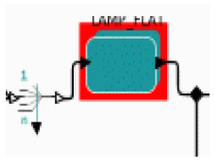


ESOReflex main features

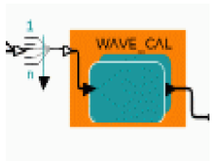


e.g. the KMOS Reflex Workflow

- 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

ESOReflex main features

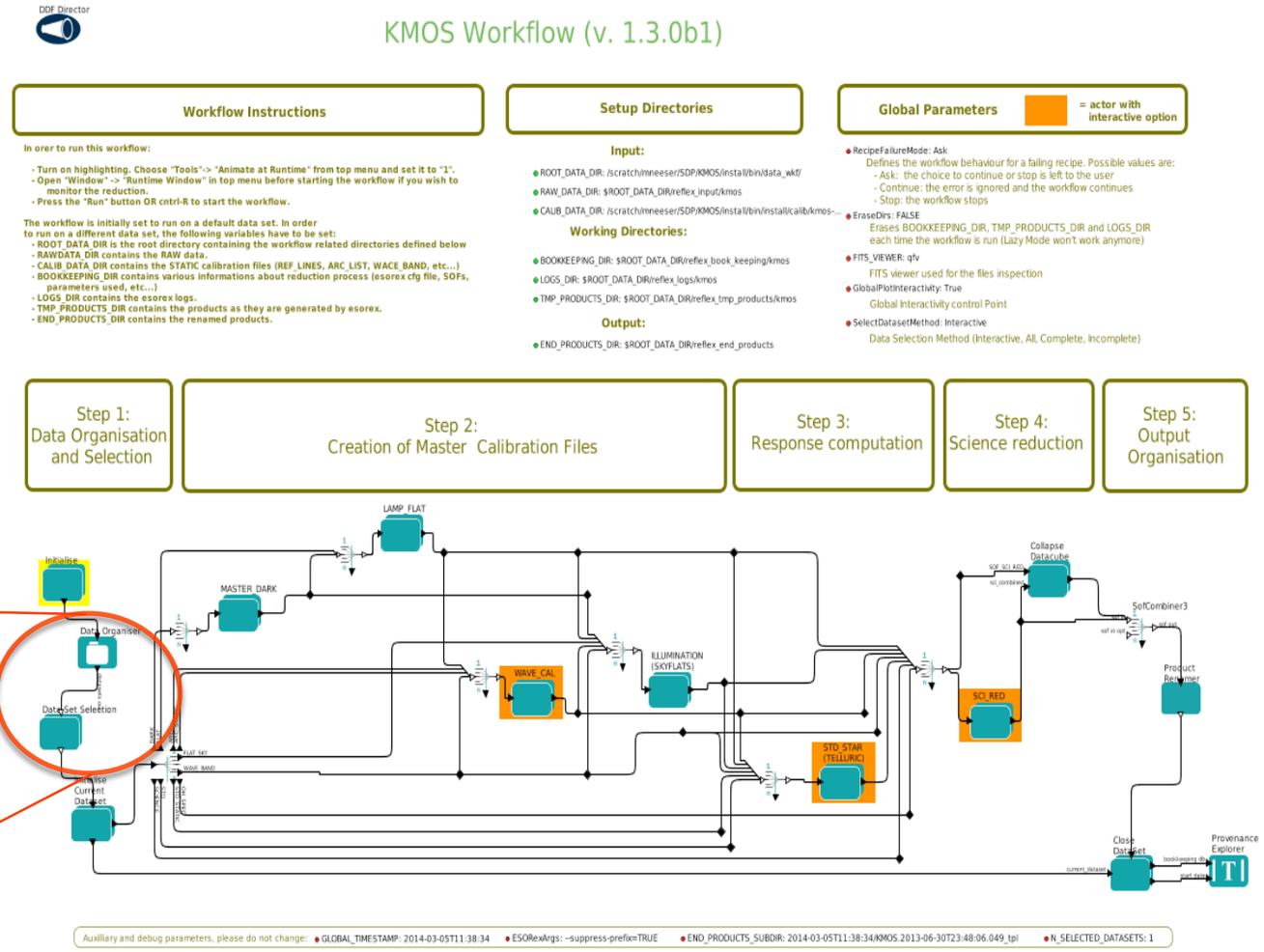
- **DataOrganiser**

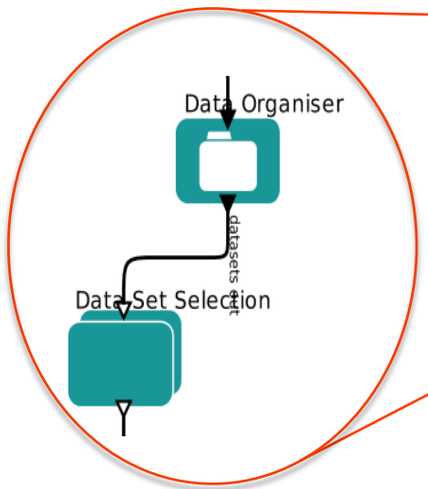
organises all input data (science & calibrations) into groups (datasets) that can be processed independently by the workflow. Each dataset has a tree structure and represents the complete calibration cascade.

The datasets are organised, classified, and associated using the data headers

==> any science frame will be optimally matched to its require 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 inspect them and select which ones to process.

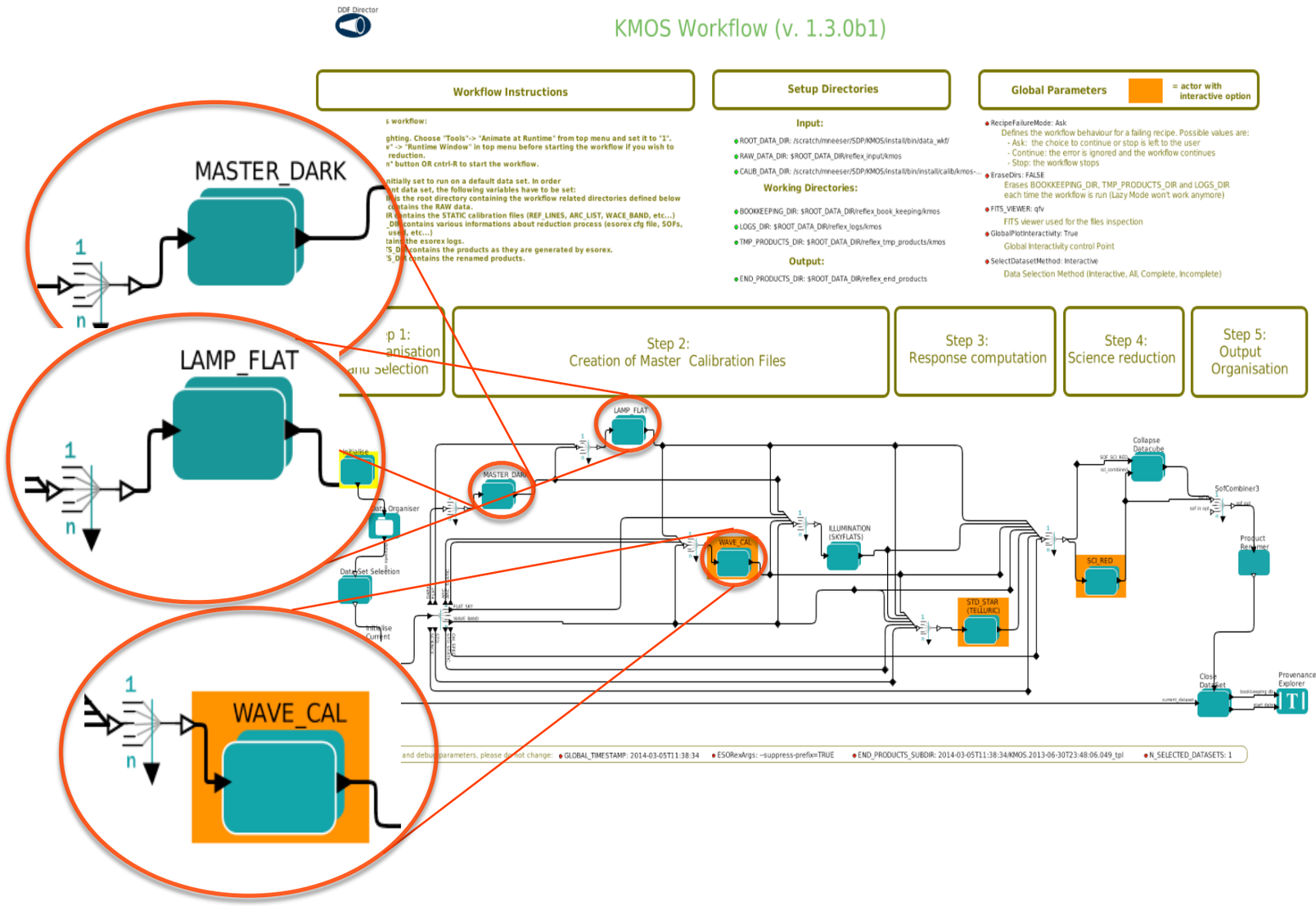
ESOReflex main features

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.



ESOReflex main features

Interactive Actors

- **Wavelength Calibration**

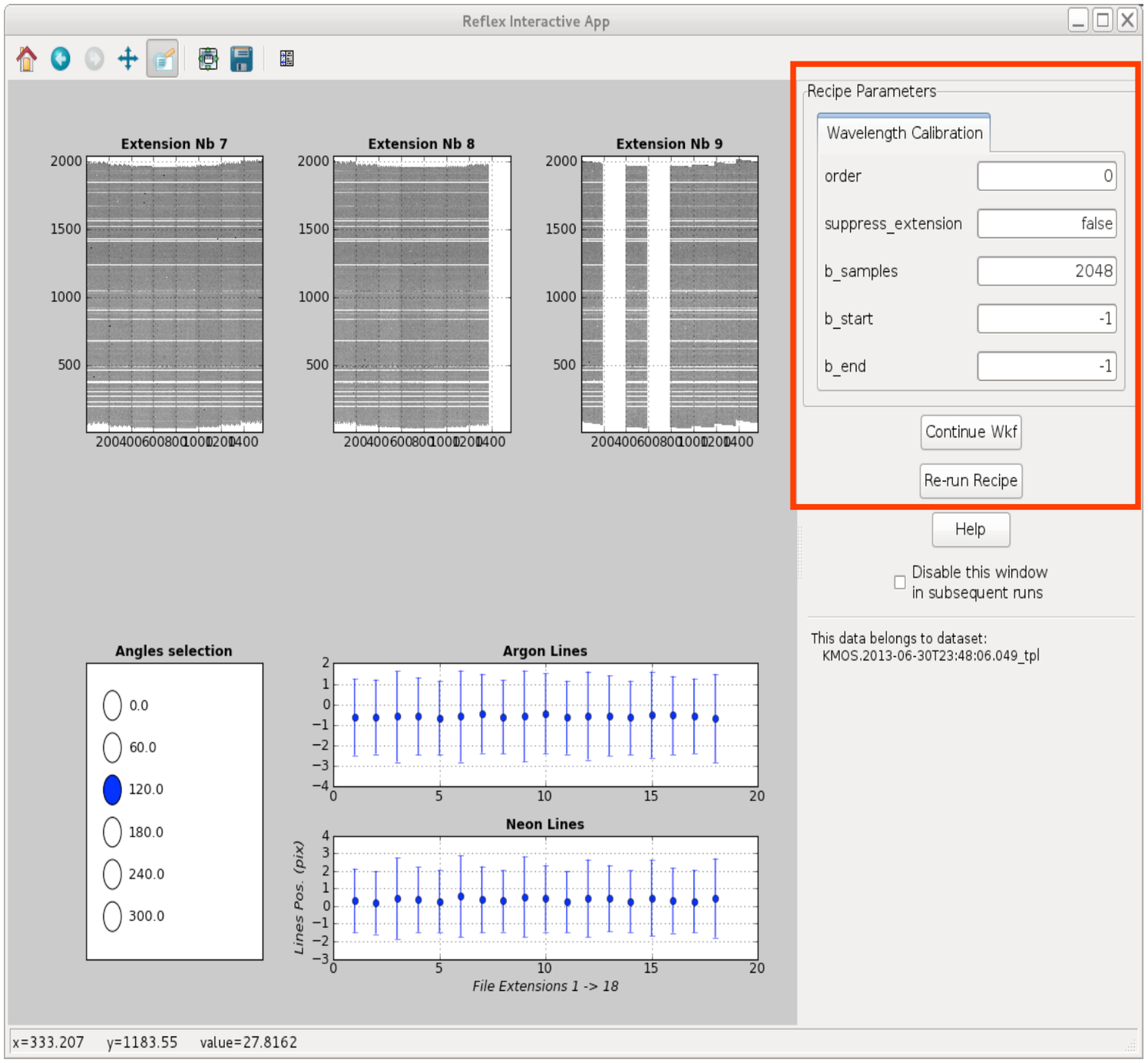
- reveals reconstructed arc lamp frames for each detector and can be selected for each of the six calibration rotator angles (0º, 60º, 120º, 180º, 240º, and 300º)

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

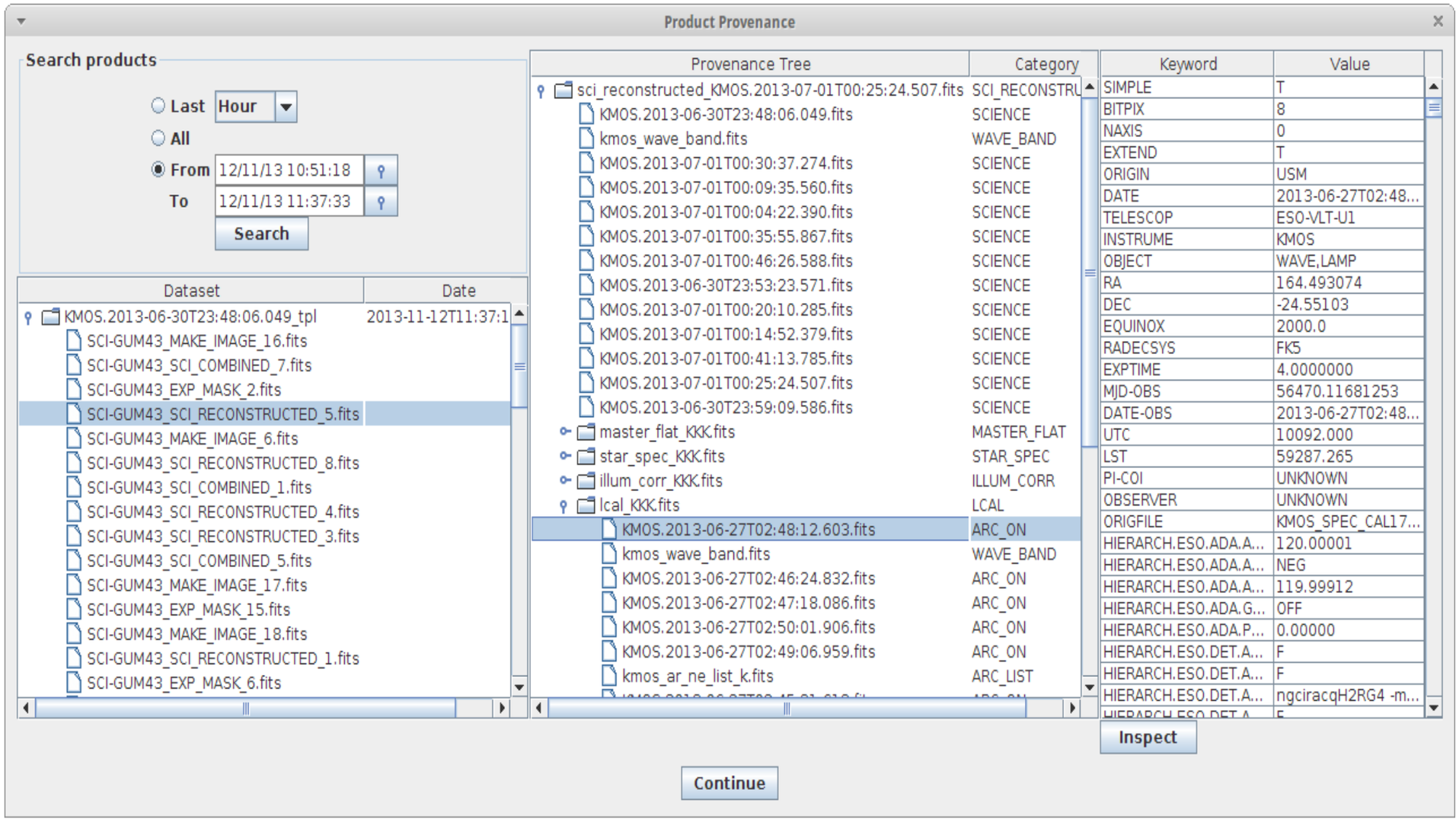
can be changed and re-run in pipeline to optimise results.



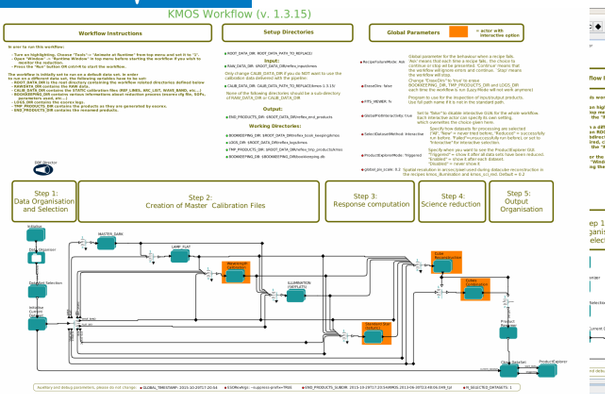
ESOReflex main features

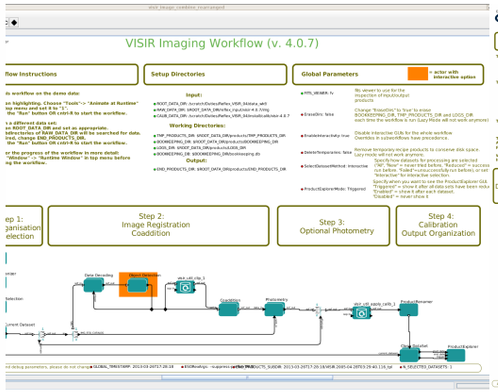
- **Product Provenance**

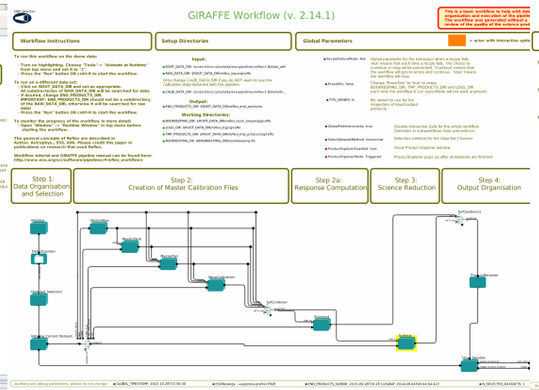
the Reflex workflow finishes with a product explorer window in which all pipeline products can be tracked to the raw & calibration frames that were used in their creation.

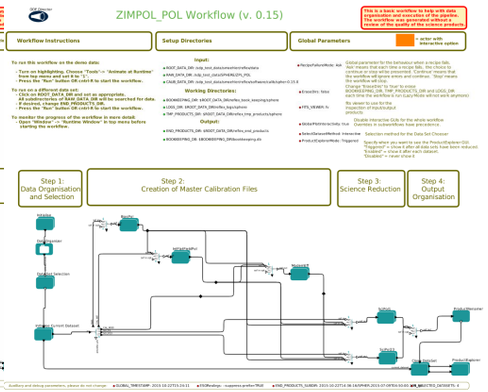


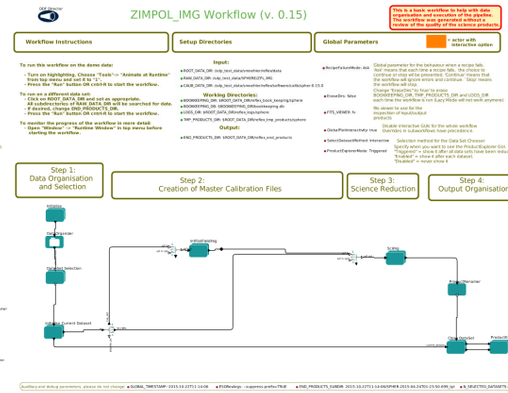
ESOReflex current workflows

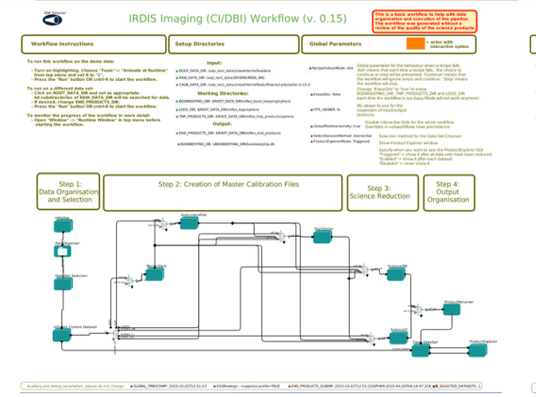


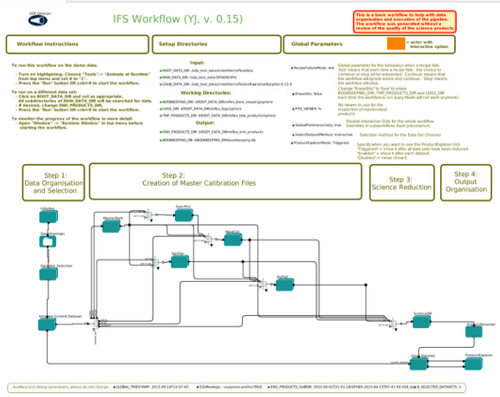


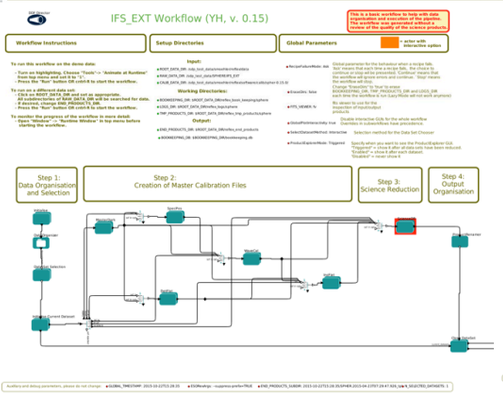


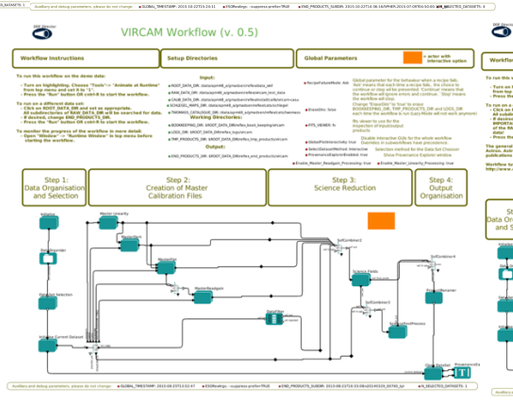


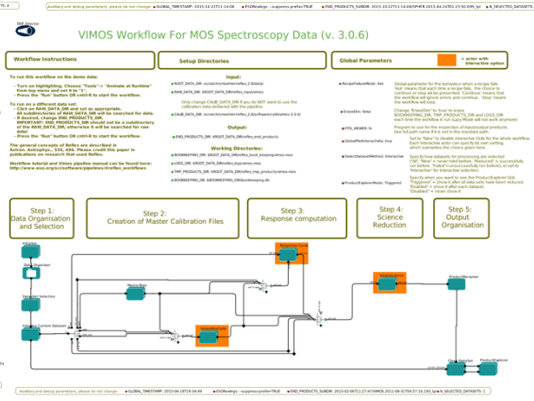


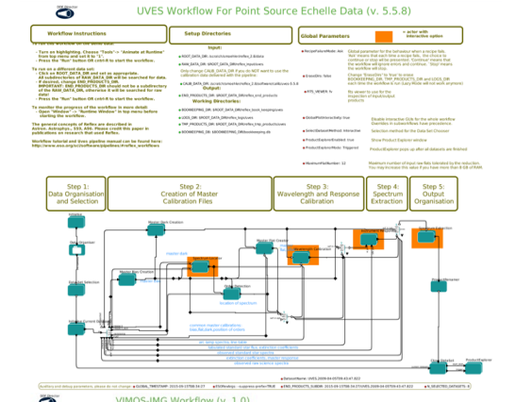


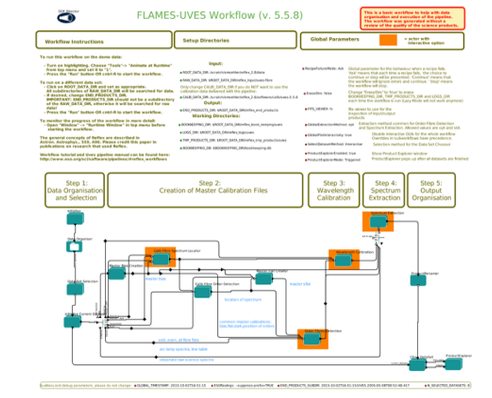


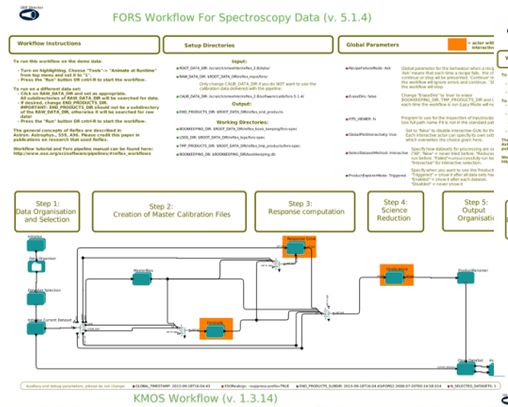


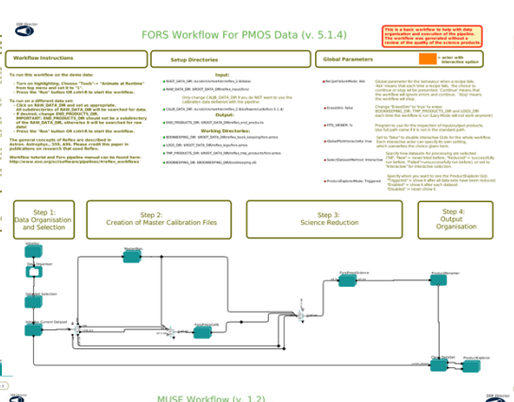


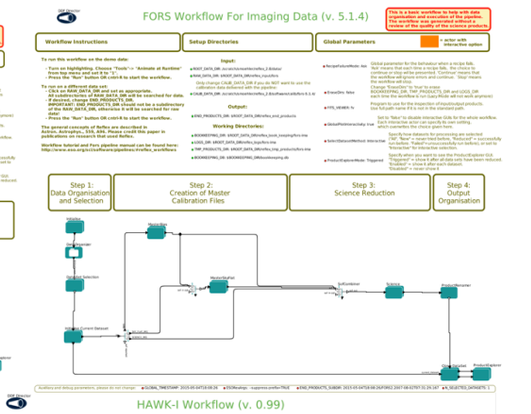


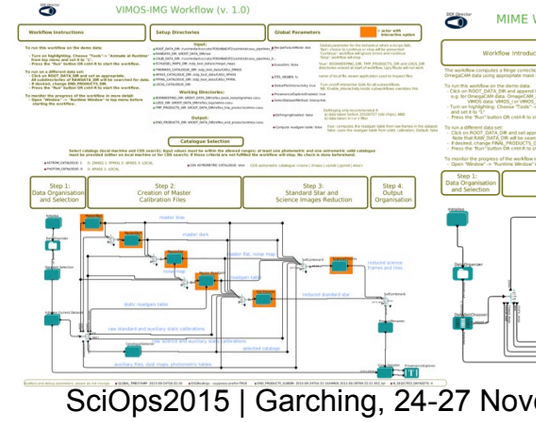


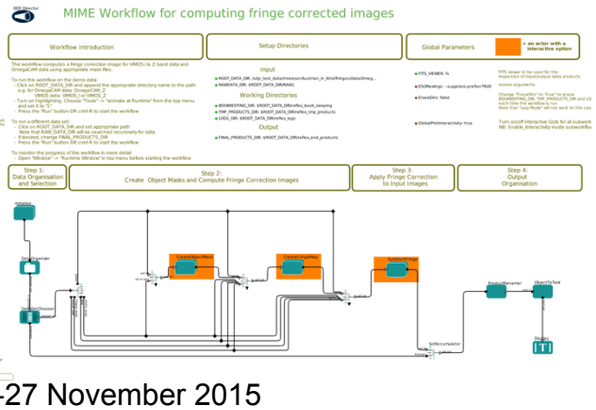


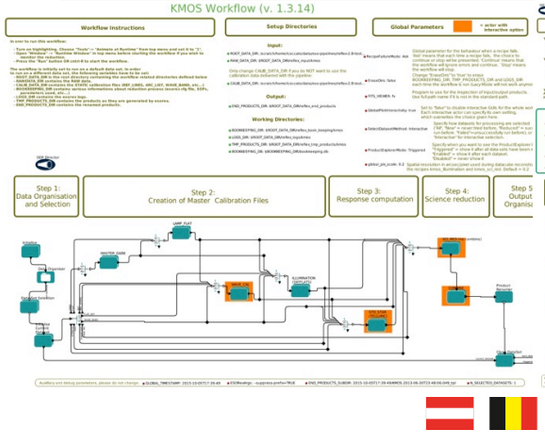


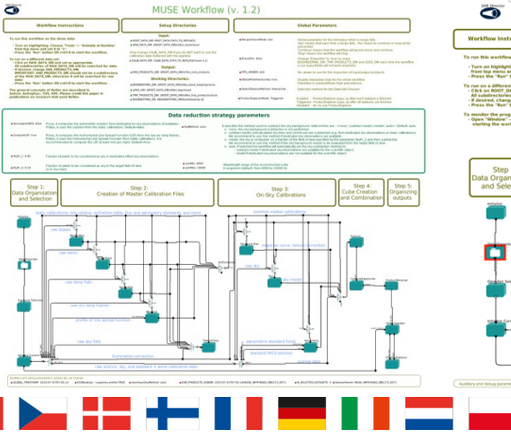


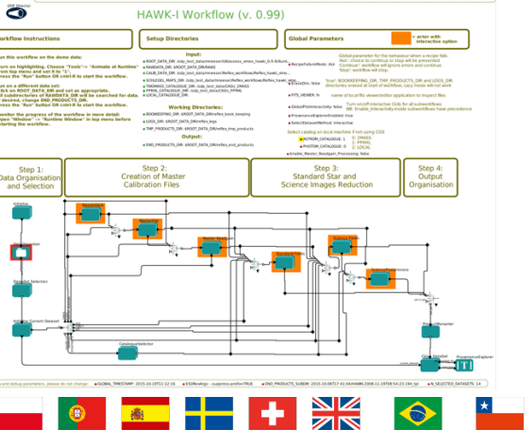












Let the work flow

ESOReflex: ESO has a system in place that allows us to deliver data reduction pipelines that are used in daily operation with an intuitive user interface.

- Current release 2.8.
- Reflex in maintenance mode.
- By early next year, about 20 workflows will be available.
- Workflows will be available for all new instruments.
- We offer workshops and tutorials on ESOReflex.