

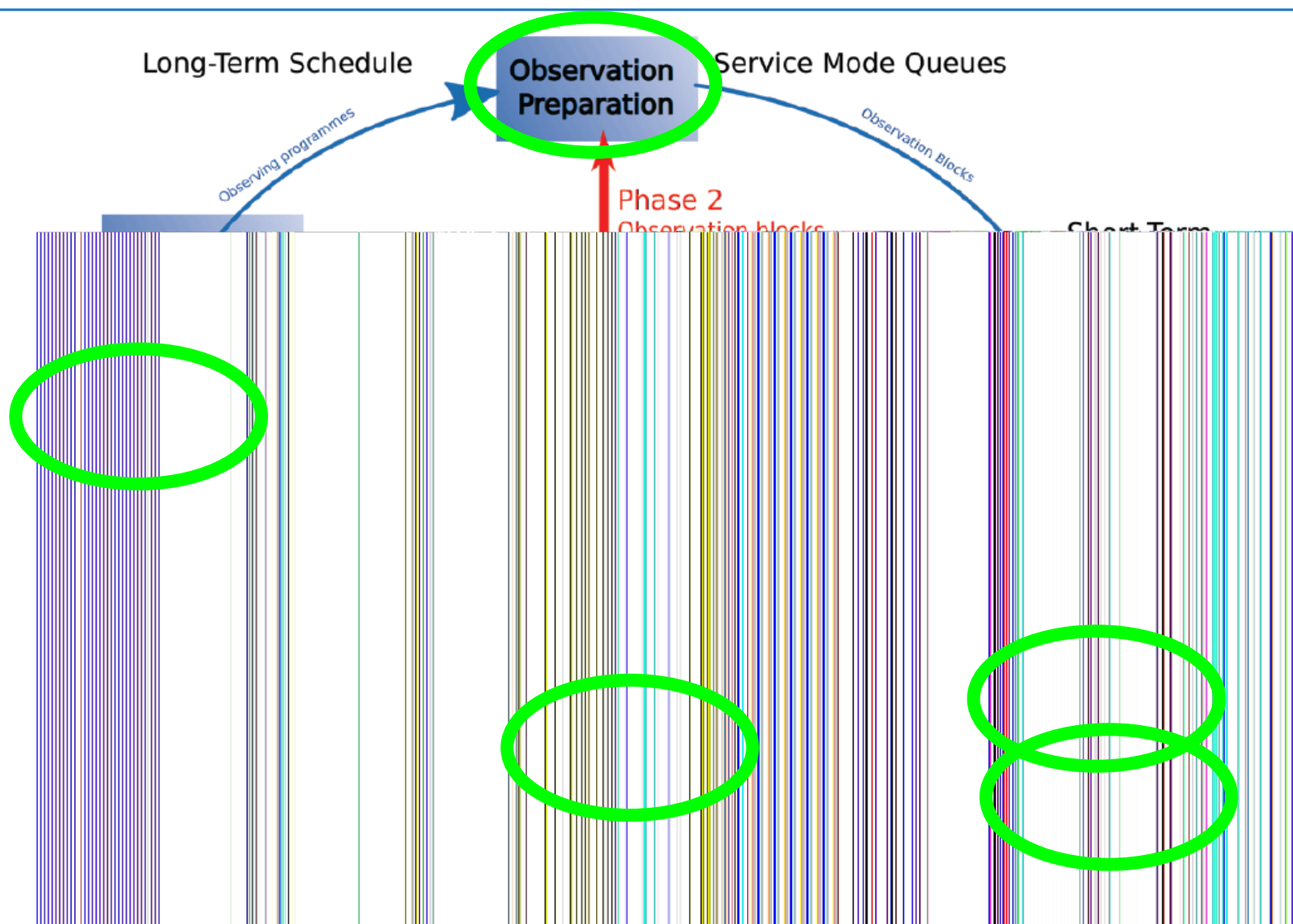


# ESO Data Reduction and Archive

Wolfram Freudling

European Southern Observatory

# The end-to-end operation model





# Usage of data reduction tools

## Observatory

- Quality control at telescope and at ESO headquarter
- Processing of pre-imaging data
- Production and certification of master calibration frames
- Production of science data products for the archive facility.

## Community

- Data reduction with custom settings of parameters
- Inspect intermediate products
- Insert customized steps into pipeline processing

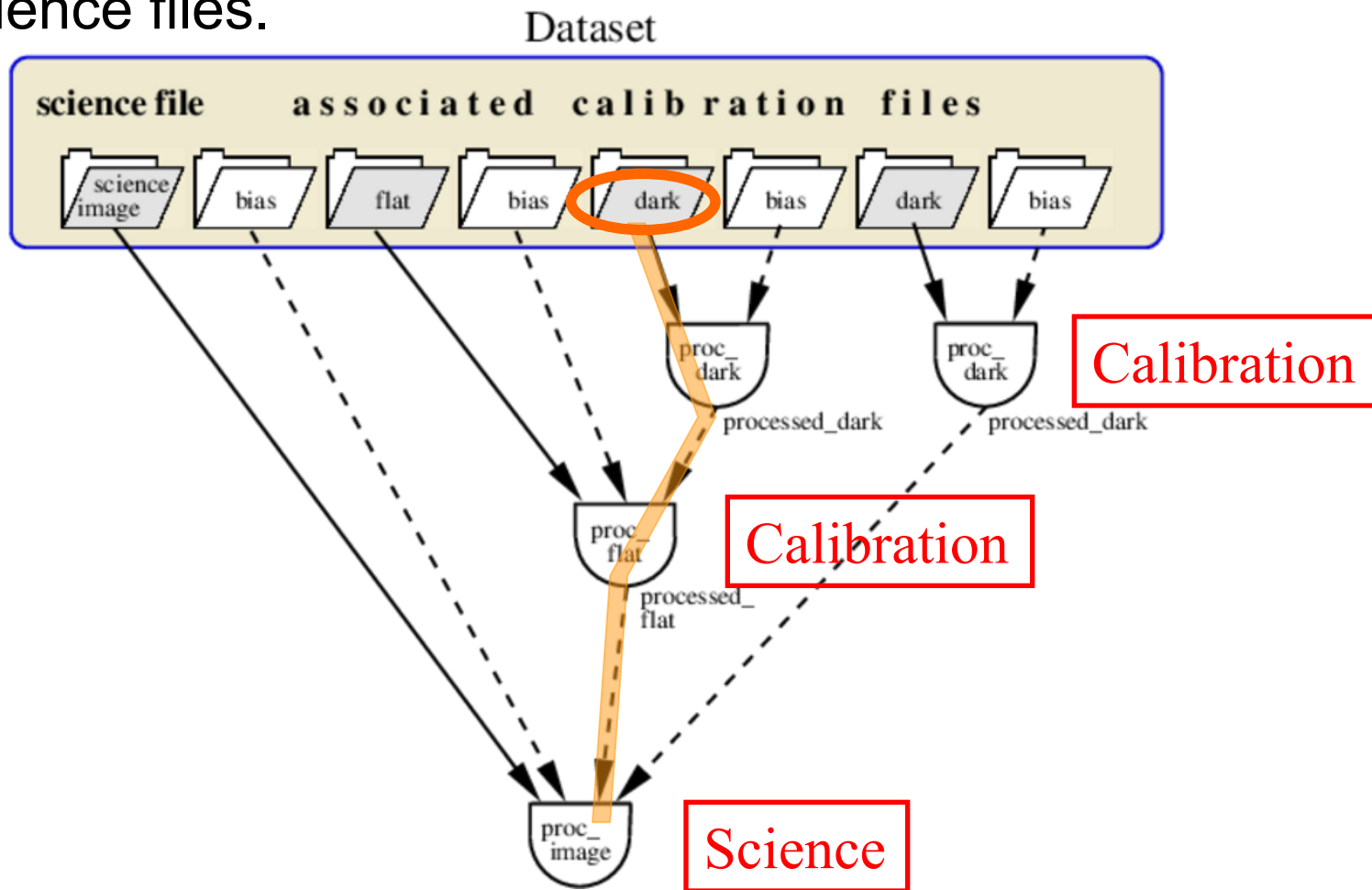


# ESO Data Reduction Infrastructure

ESO Internal only	Public	communalities
Data organization		
DO at telescope, ABbuilder QC Garching	CalSelector service, Gasgano, Reflex DO	OCA language
Algorithms		
Quick look recipes	CPL pipeline recipes, Reflex Python plotting, (Molecfite and other tools)	CPL
Data Reduction workflows		
Paranal & QC cascades, Phoenix	Reflex	Esorex recipe execution

# Data Organization

Organise files into data sets = all files needed to reduce one set of science files.





# Data Organisation

- Pipeline (ESO Observatory, Data Organiser)
  - On-the-fly data processing (event driven)
  - Template-based processing
  - Static calibration database (only certified products are used)
- Quality Control (ESO Headquarters, ABbuilder)
  - Batch processing of complete data sets (all science and calibration data produced by one ESO instrument in one night)
  - Best available calibrations are used => data must be organized according to the **Calibration Cascade**
- Science Archive (CalSelector)
  - Web-based association of calibration data in archive (raw and reduced) to science data
- Data Reduction by Community (Reflex, Gasgano)
  - Data organisation of files on disk for different processing cascades.



# OCA Rules

- Data organisation defined as text files “OCA rules”
- Three types of rules:

➤ Classification ( „This is a **Raw Dark**“)

```
if DPR.CATG=="CALIB" and DPR.TYPE=="DARK" then
{
  RAW.TYPE = "DARK";
}
```

➤ Organization („These Raw Darks are processed together“)

```
select execute(MAKEDARK) from inputFiles
where RAW.TYPE=="DARK" group by TPL.START
```

➤ Association (“select Biases based on properties of Raw Darks“  
or “process these Raw Darks together using selected Biases“)

```
action MAKEDARK
{
  select file as MASTER_BIAS from calibFiles where
  PRO.CATG=="MASTER_BIAS" and inputFile.DET.WIN1.BINX==DET.WIN1.BINX;
}
```



# CPL Pipelines

- Consortia are required to deliver pipeline recipes coded in C using the “ESO Common Pipeline Library” and “High

Instrument	Release Notes	Package	User Manual	Cookbook	Additional Documents	Additional Datasets	EsoReflex	Status
AMBER	<a href="#">2015-03-06</a>	<a href="#">4.3.3</a>	<a href="#">4.3.2</a>					Operational on hold
CRIRES	<a href="#">2015-08-04</a>	<a href="#">2.3.3</a>	<a href="#">1.13</a>	<a href="#">Cookbook</a>				Operational on hold
EFOSC	<a href="#">2015-07-10</a>	<a href="#">2.2.4</a>	<a href="#">1.0</a>			<a href="#">Demo Data</a>		End of maintenance
FORS	<a href="#">2015-09-18</a>	<a href="#">5.1.4</a>	<a href="#">5.2</a>			<a href="#">Demo Data (29 MB)</a>	Tutorial: <a href="#">1.0 (FORS-IMG)</a> Tutorial: <a href="#">1.0 (FORS-PMOS)</a> Tutorial: <a href="#">1.9 (FORS-SPEC)</a> Demo Data: <a href="#">0.8</a>	Active
GIRAFFE	<a href="#">2015-11-02</a>	<a href="#">2.14.2</a>	<a href="#">2.14.2</a>	<a href="#">Cookbook</a>		<a href="#">Standard Calibration Files page</a>		Operational on hold
HAWKI	<a href="#">2015-04-20</a>	<a href="#">1.8.18</a>	<a href="#">1.11</a>			<a href="#">Demonstration Package (2.5 GB)</a>		Operational on hold
ISAAC	<a href="#">2015-04-17</a>	<a href="#">6.1.5</a>	<a href="#">1.4</a>			<a href="#">Static Calibration Files (50 MB)</a>		End of maintenance
KMOS	<a href="#">2016-01-22</a>	<a href="#">1.3.17</a>	<a href="#">2.17</a>				Tutorial: <a href="#">1.6</a> Demo Data: <a href="#">1.2</a>	Active
MIDI	<a href="#">2015-04-15</a>	<a href="#">2.8.4</a>	<a href="#">2.8.3</a>					End of maintenance
MUSE	<a href="#">2015-10-06</a>	<a href="#">1.2.1</a>	<a href="#">1.2.1</a>			<a href="#">MUSE IFU 6 trace tables</a> <a href="#">Legacy MUSE static calibrations</a>	Tutorial: <a href="#">7.0</a> Demo Data: <a href="#">1.3</a>	Active
NACO	<a href="#">2015-06-01</a>	<a href="#">4.4.1</a>	<a href="#">1.1</a>					Operational on hold
SINFONI	<a href="#">2015-10-26</a>	<a href="#">2.7.0</a>	<a href="#">19.5</a>		<a href="#">ADA IV 2006 paper</a>	<a href="#">Calibration Database Example (255 MB)</a> <a href="#">Demonstration Package (1.2 GB)</a>	Tutorial: <a href="#">1.5</a> Demo Data: <a href="#">0.2</a>	Operational on hold
SOFI	<a href="#">2015-04-17</a>	<a href="#">1.5.6</a>	<a href="#">1.2</a>					End of maintenance
SPHERE	<a href="#">2015-03-10</a>	<a href="#">0.15.0</a>						Active
UVES	<a href="#">2015-09-14</a>	<a href="#">5.5.7</a>	<a href="#">22.11 (UVES)</a> <a href="#">18.5 (UVES-FIBRE)</a>			<a href="#">Demonstration Package (2.0 GB)</a>	Tutorial: <a href="#">6.6 (UVES)</a> Tutorial: <a href="#">1.5 (UVES-FIBRE)</a> Demo Data: <a href="#">4.4</a>	Operational on hold
VIMOS	<a href="#">2015-10-05</a>	<a href="#">3.0.6</a>	<a href="#">7.0</a>			<a href="#">Demonstration Package (1.7 GB)</a>	Tutorial: <a href="#">2.3 (VIMOS-IFU)</a> Tutorial: <a href="#">2.0 (VIMOS-MOS)</a> Demo Data: <a href="#">0.4</a>	Active
VISIR	<a href="#">2016-02-25</a>	<a href="#">4.1.7</a>	<a href="#">1.5</a>				Demo Data: <a href="#">0.1</a>	Operational on hold
XSHOOTER	<a href="#">2015-09-14</a>	<a href="#">2.6.8</a>	<a href="#">12.7</a>			<a href="#">Additional NIR telluric model catalog (190 MB)</a>	Tutorial: <a href="#">2.6</a> Demo Data: <a href="#">1.2</a>	Operational on hold

- Scientific oversight by Garching instrument project scientists during development, Science Data Products Group and Paranal instrument scientist during operation





# High level Data Reduction Library HDRL

A. Gabasch

- Overscan computation and subtraction
- Master frame combination(Bias/Dark/Flat)
- Cosmic ray detection on a single image
- Bad pixel determination on
  - Single images.
  - Stacks of identical images, e.g. bias/dark frames.
  - Sequence of images, e.g. domeflats with different exposure time.
- Computation of the Strehl ratio
- Fringe detection and removal
- Source detection/extraction

More to come:

- Optimal Extraction
- Wavelength calibration
- Detection of pick-up noise
- ...

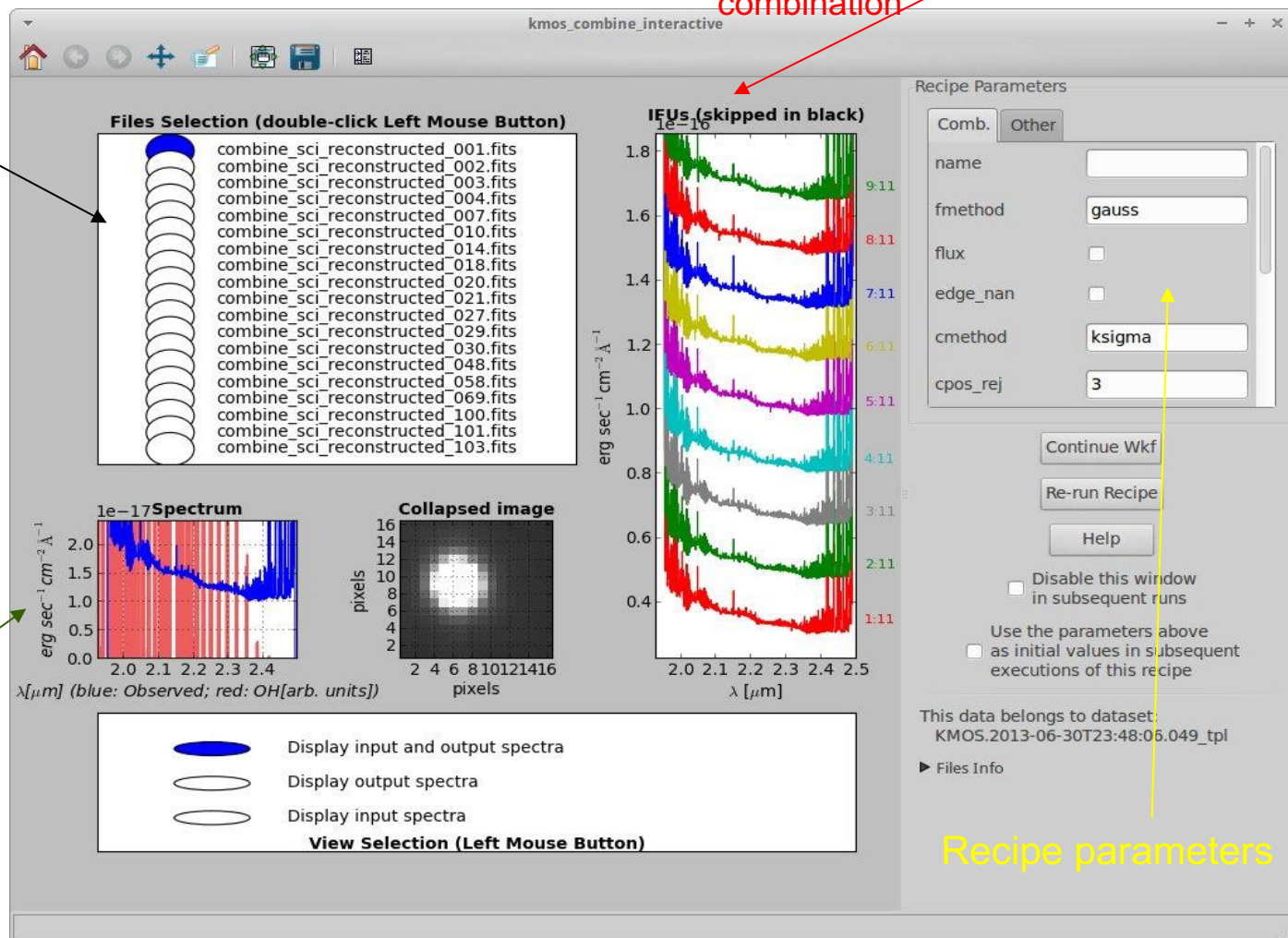


# KMOS: combining reconstructed cubes

List of combined datacubes  
for each object

Representative spectrum and  
image of the selected object  
(from final combined cube)

Representative spectra of individual datacubes  
that are combined for the selected object. The  
user can (de)select the exposures from final  
combination




Recipe parameters



# Data Reduction Workflows

- Even when algorithms (recipes) are given, data reduction involves many decisions:
  - Which data do I process first?
  - Which files are processed by which recipes?
  - Which files need to be re-processed?
  - How do I organize the output?
- Workflow differs for different use cases. Example:
  - QC0 use static calibration
  - QC1 processes all calibrations independent of later use
  - Desktop reduction processes all data for given datasets

# Scientific Workflow Systems

- System to define and execute series of data manipulation steps
- Flexible intuitive workflow essential for desktop data reduction
- **Reflex** uses Kepler  <https://kepler-project.org>
- Kepler provides the graphical user interface
- **Reflex** is a collection of Kepler components (“actors”) that allow to execute ESO recipes, data display GUIs, and Python scripts
- Available for Linux and OS X

## Workflow Instructions

To run this workflow on the demo data:

- Turn on highlighting. Choose "Tools"-> "Animate at Runtime" from top menu and set it to "1".
- Press the "Run" button OR cntrl-R to start the workflow.

To run on a different data set:

- Click on ROOT\_DATA\_DIR and set as appropriate.
- All subdirectories of RAWDATA\_DIR will be searched for data.
- If desired, change END\_PRODUCTS\_DIR.
- Press the "Run" button OR cntrl-R to start the workflow.

To monitor the progress of the workflow in more detail:

- Open "Window" -> "Runtime Window" in top menu before starting the workflow.

## Setup Directories

## Input:

- ROOT\_DATA\_DIR: /sdp\_test\_data/mneiser/UK/assess\_hawki\_1.1/A1689
- RAWDATA\_DIR: \$ROOT\_DATA\_DIR/RAW
- CALIB\_DATA\_DIR: /sdp\_test\_data/mneiser/Reflex\_workflows/Reflex\_hawki\_vimo...

## Working Directories:

- TWOMASS\_CATALOGUE\_DIR: /sdp\_test\_data/CASU\_2MASS
- PPMXL\_CATALOGUE\_DIR: /sdp\_test\_data/CASU\_PPMXL
- LOCAL\_CATALOGUE\_DIR:
- BOOKKEEPING\_DIR: \$ROOT\_DATA\_DIR/reflex\_book\_keeping/hawki
- LOGS\_DIR: \$ROOT\_DATA\_DIR/reflex\_logs/hawki
- TMP\_PRODUCTS\_DIR: \$ROOT\_DATA\_DIR/reflex\_tmp\_products/hawki

## Output:

- END\_PRODUCTS\_DIR: \$ROOT\_DATA\_DIR/reflex\_end\_products

## Global Parameters

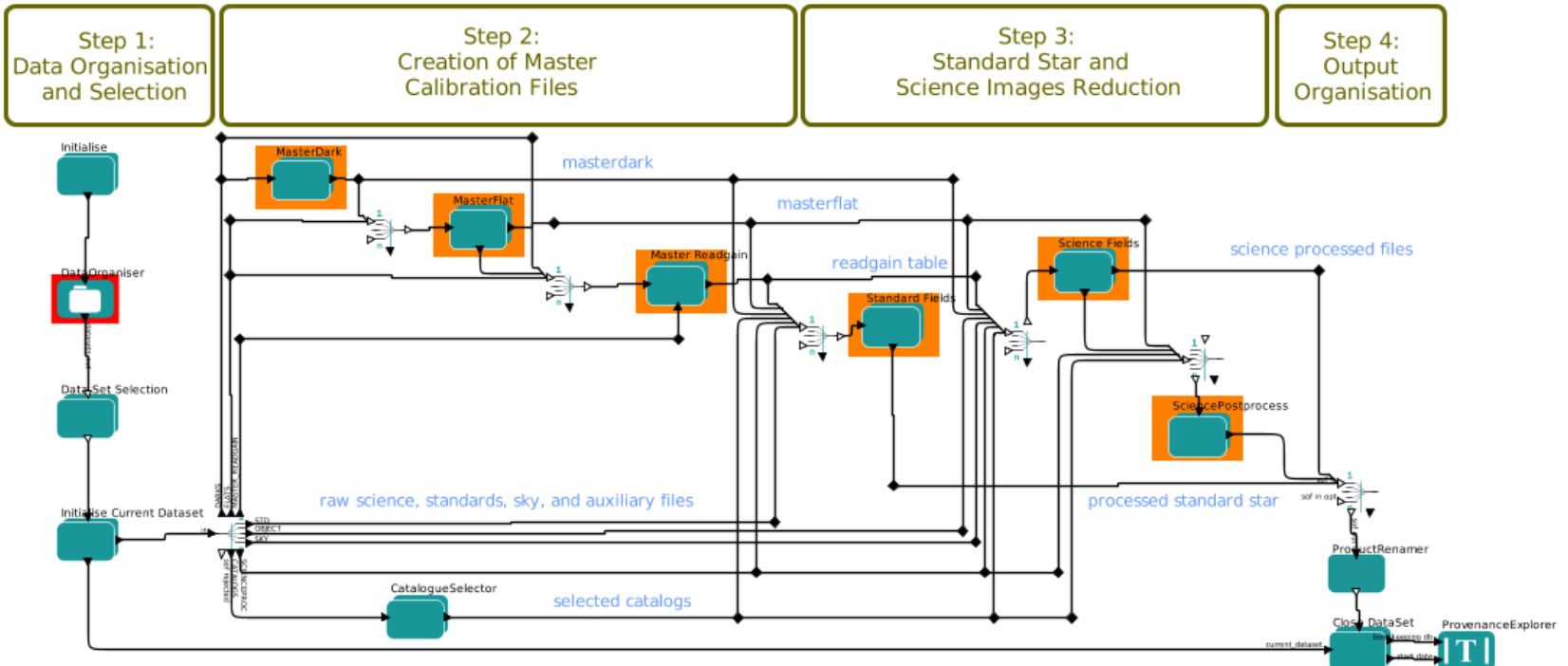
 = actor with interactive option

- RecipeFailureMode: Ask Global parameter for the behaviour when a recipe fails. 'Ask': choice to continue or stop will be presented 'Continue': workflow will ignore errors and continue 'Stop': workflow will stop.
- EraseDirs: false 'true': BOOKKEEPING\_DIR, TMP\_PRODUCTS\_DIR and LOGS\_DIR directories erased at start of workflow; Lazy mode will not work
- FITS\_VIEWER: fv name of local fits viewer/editor application to inspect files
- GlobalPlotInteractivity: true Turn on/off interactive GUIs for all subworkflows NB: Enable\_Interactivity inside subworkflows have precedence
- ProvenanceExplorerEnabled: true
- SelectDatasetMethod: interactive
- Compute\_readgain\_table: false true: computed the readgain table from raw frames in the dataset. false: uses the readgain table from static calibrations. Default: false.

## Catalogue Selection

Select catalogues (on local machine and/or for CDS search). Input values must be within the allowed ranges; at least one photometric and one astrometric valid catalogues must be provided (either from local machine or for CDS search). If these criteria are not fulfilled, the workflow will stop. No check is done beforehand.

- ASTROM\_CATALOGUE: 1 Select catalogues on local machine
- CDS ASTROMETRIC CATALOGUE: wise CDS astrometric catalogue, <none | 2mass | usnob | ppmxl | wise>
- PHOTOM\_CATALOGUE: 0 0: 2MASS; 1: PPMXL; 2: LOCAL
- CDS PHOTOMETRIC CATALOGUE: 2mass CDS photometric catalogue, <none | 2mass | ppmxl | wise>



# KMOS WORKFLOW

## Global parameters for data reduction

• global\_pix\_scale: 0.2

Spatial resolution in arcsec/pixel used during datacube reconstruction in the recipes kmos\_illumination and kmos\_sci\_red. Default = 0.2

• UseSkyFlats?: no

Specify the strategy for illumination correction. If set to yes, the twilight sky flats will be used if present. If set to yes but no sky flats are present in the dataset, the illumination correction is not done. If set to no, the internal lamp flats are used for illumination correction. Default value: yes.

• telluric and response correction: 0

Specify the response and telluric correction to use. Default=0. Valid entries:

0: Response, telluric correction, and zeropoint evaluated from standard star or user provided file (category: TELLURIC). User provided file has priority. If none are present, the average response curve and zeropoint from static calibration is applied if present, otherwise products are not flux calibrated.

1: Apply only telluric correction and zeropoint from user-provided file (category TELLURIC\_CORR).

2: Apply only response correction from static calibration (category: RESPONSE). Zeropoint is computed either from standard star or from user-provided file (category: TELLURIC). This option is recommended when correcting for telluric absorption with external tools (e.g., molecfit).

3: Apply response correction from static calibration (category: RESPONSE) and telluric correction from user provided file (category: TELLURIC\_CORR). Zeropoint from user-provided file is used.

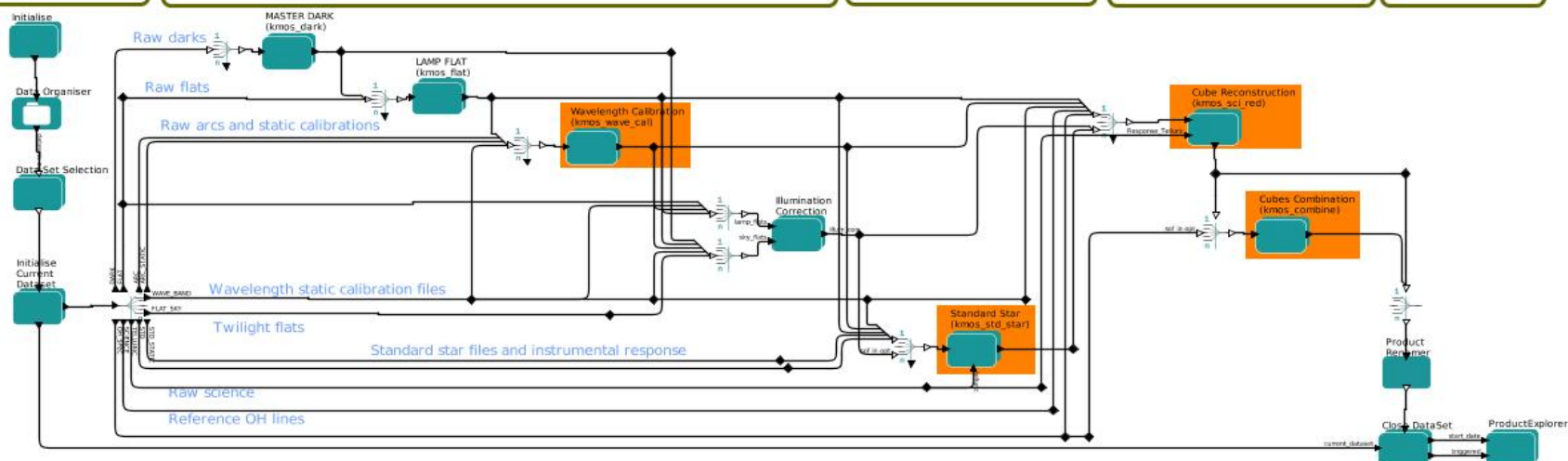
Step 1:  
Data Organisation  
and Selection

Step 2:  
Creation of Master Calibration Files

Step 3:  
Response computation  
Telluric correction

Step 4:  
Science reduction

Step 5:  
Output  
Organisation



Global parameters that allows to define:

- The illumination correction strategy: twilight flats or lamp flats.
- The Telluric correction and response curve correction strategy:
  - Response and telluric correction from pipeline (using observed standard star)
  - Merge the static Response calibration with an user-provided telluric correction (obtained, e.g., using molecfit).
  - Response correction only (from static calibration).
  - Telluric correction only (from user-provided calibration)





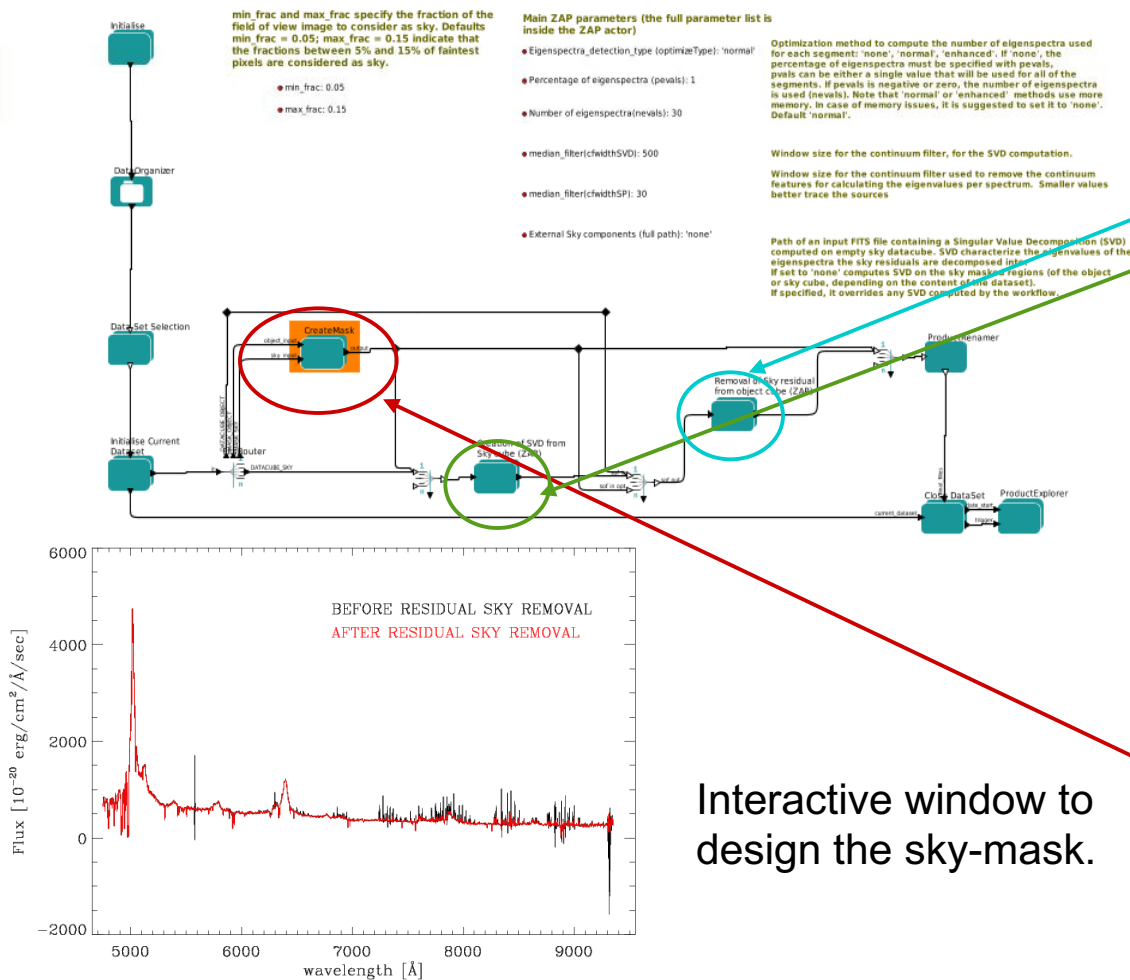
# MUSE-ZAP workflow

Step 1:  
Data Organisation  
and Selection

Step 2:  
Creation of Sky Mask

Step 3:  
Removing residual sky lines  
from datacube

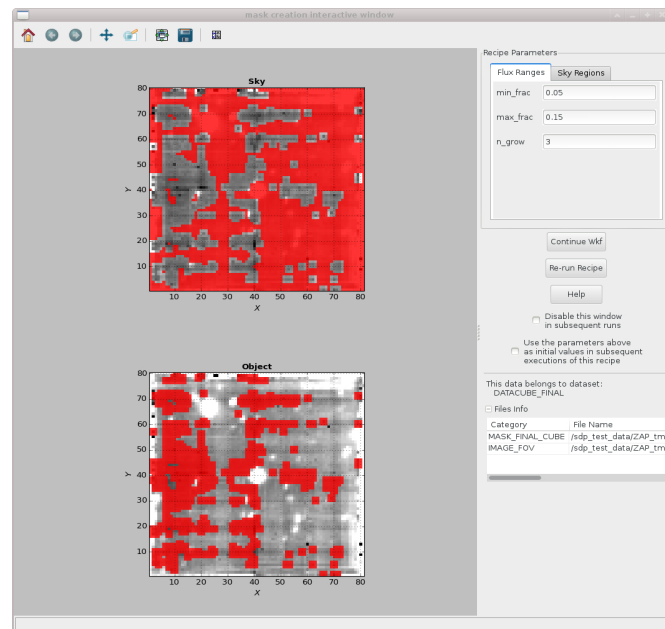
Step 4:  
Output Organisation



Sky residuals are decomposed into principal components either using:

- Sky regions around the target;
- Dedicated sky observations (if present in the dataset).

Interactive window to design the sky-mask.



- It uses the Python code by Soto et al. ([2016MNRAS.458.3210S](#))
- It is independent of the muse pipeline.



# Science Archive

- 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**)
  - ~700 TB of data in ~40 million files and ~30 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





# Data products in the SAF

**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, MUSE, then HAWK-I and VIMOS IMG, KMOS, ...
  - Migration 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

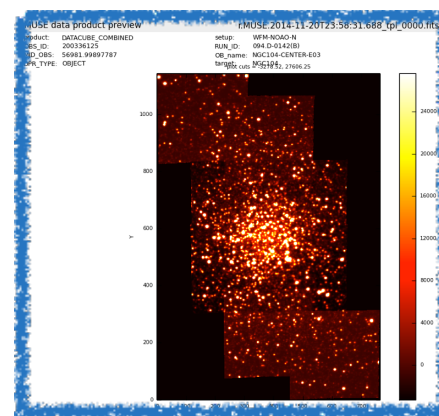
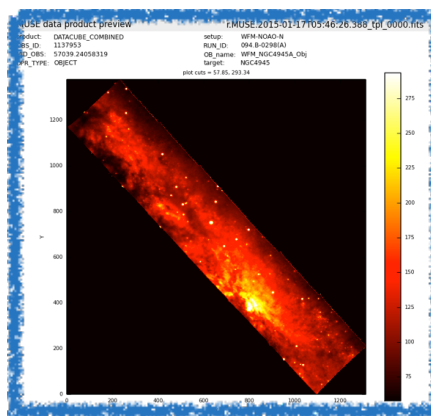
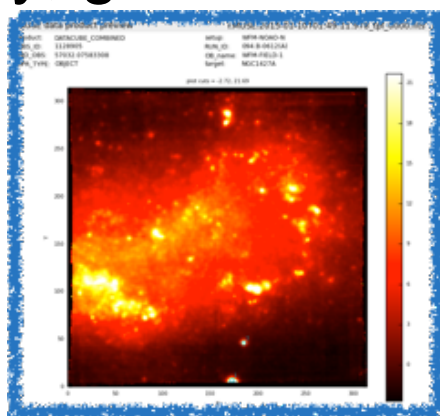
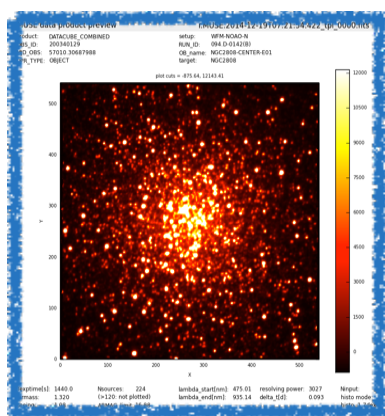
Arnaboldi et al: Summary in December 2013 Messenger: papers from ESO, the Survey Teams and archive user





# Latest addition: MUSE

- We have just published processed data for the entire data history, generated in-house



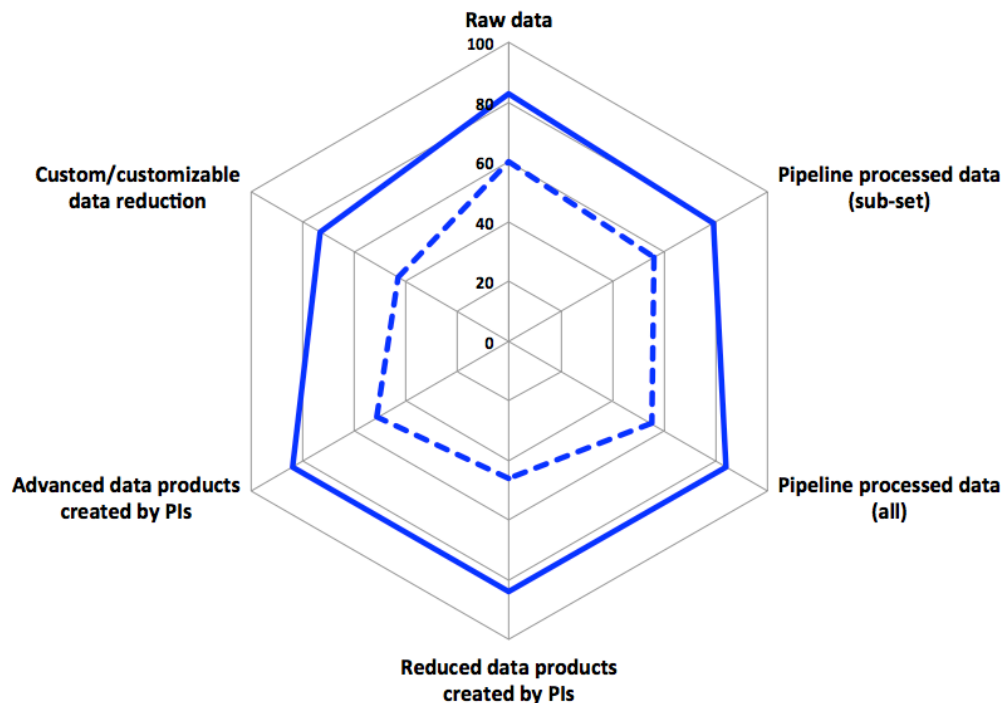
Following talk by Reinhard Hanuschik



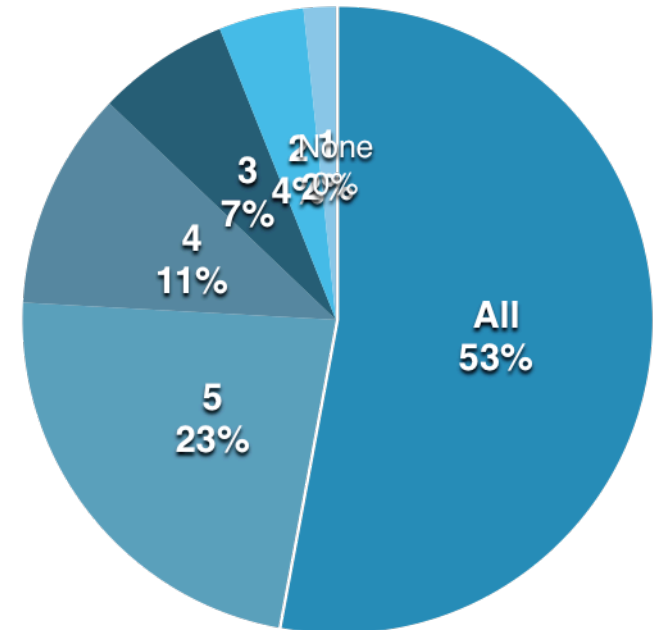
# Polling the community: ESO2020 questionnaire

- “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

— Somewhat important to Essential — Essential or Very important



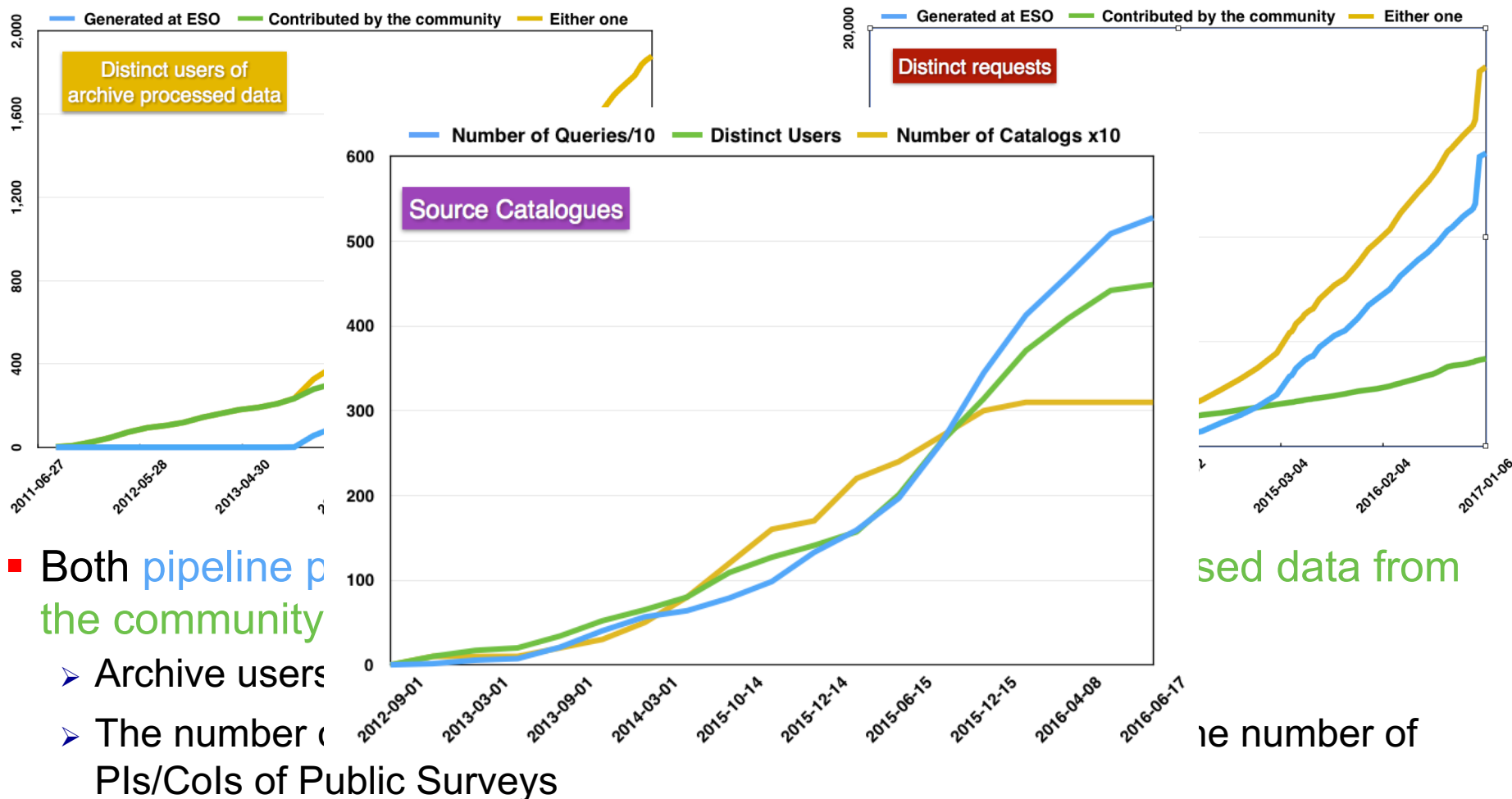
Number of archive data categories considered  
“Somewhat important” to “Essential”



# Archive access to data: building a community



Romaniello et al, The Messenger, March 2016



- Both pipeline and data from the community

➤ Archive users

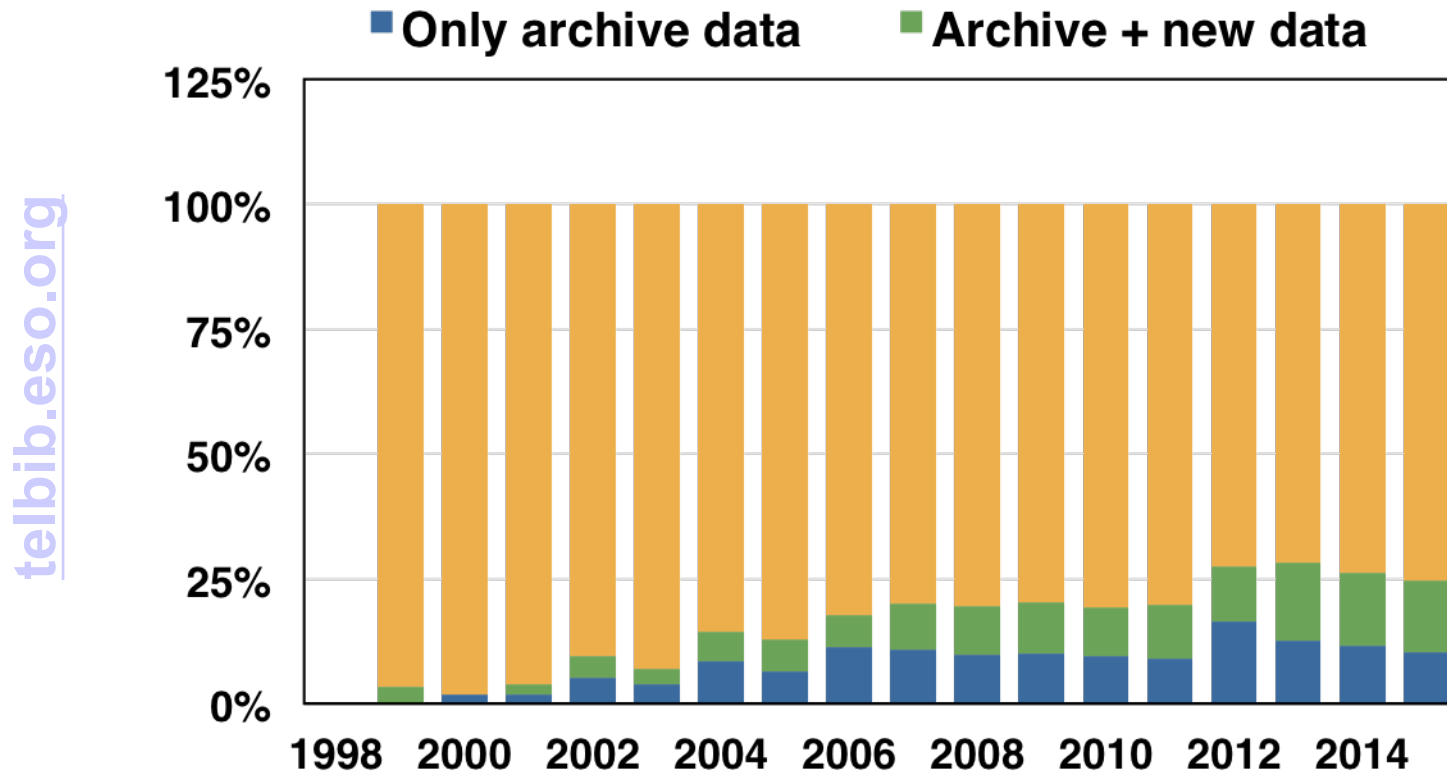
➤ The number of

PIs/Cols of Public Surveys

- Availability of products didn't decrease access to the corresponding raw data

# The science return of the SAF

- Archive refereed papers
  - ~25% of ESO's output
  - ~5% of La Silla Paranal data only published as archive papers





# Archive services: status quo

- **The problem:** as a user, how do I find my way among 40 million files described by 30 billion keywords?
  - Classical form based interface
  - Selection of calibration data (raw or processed)
- The science potential of ESO's Science Archive Facility is limited by its limited search capabilities
  - 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
- Especially true 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 archive, 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 science keywords that 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



# Some Ongoing Projects

- **Data Organization**
  - Upgrade of OCA.
  - Will allow more flexible design of pipeline recipes
  - Support for more complex conditions
- **Pipelines**
  - Support for new instruments
  - Use Molecfit in pipelines (KMOS)
  - Additions to HDRL to increase communalities among pipelines
- **ESO Archive Facility**
  - Expand holding of reduced data to new instruments: Hawk-I, VIMOS imaging, KMOS
  - Interface





# Summary

- ESO supports data reductions for 3 different use cases:
  - Operations, quality control
  - Archive
  - Community
- Different use cases use different strategies and tools for
  - Data organization
  - Reduction workflow
- Different use cases use a common underlying infrastructure
  - Pipeline recipes
  - OCA data organisation