Delivering Data Reduction Pipelines to Science Users

Wolfram Freudling and Martino Romaniello

European Southern Observatory, Karl-Schwarzschild-Str 2, 85748 Garching, Germany

ABSTRACT

The European Southern Observatory has a long history of providing specialized algorithms called recipes for each of its instruments. However, calling these recipes in sequence and providing the proper input to each of the recipes is a challenging and time consuming effort. The efficiency of data reduction can vastly be improved by using automatic workflows to organise data and execute a sequence of data reduction steps. To realize such efficiency gains, we designed a system called ESOREFLEX that allows intuitive representation, execution and modification of the data reduction workflow, and has facilities for inspection and interaction with the data. ESOREFLEX uses a number of innovative concepts and has been described in Ref. 1. In October 2015, the complete system was released to the public. It includes systems for automatic data organization and visualization, interaction with recipes, and the exploration of the provenance tree of intermediate and final data products. ESOREFLEX is highly efficient, using its internal bookkeeping database to recognize and skip previously completed steps during repeated processing of the same or similar data sets.

Keywords: Manuscript format, template, SPIE Proceedings, LaTeX

1. INTRODUCTION

2. 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

3. 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

4. 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[2] 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

5. ESOREFLEX

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

6. 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

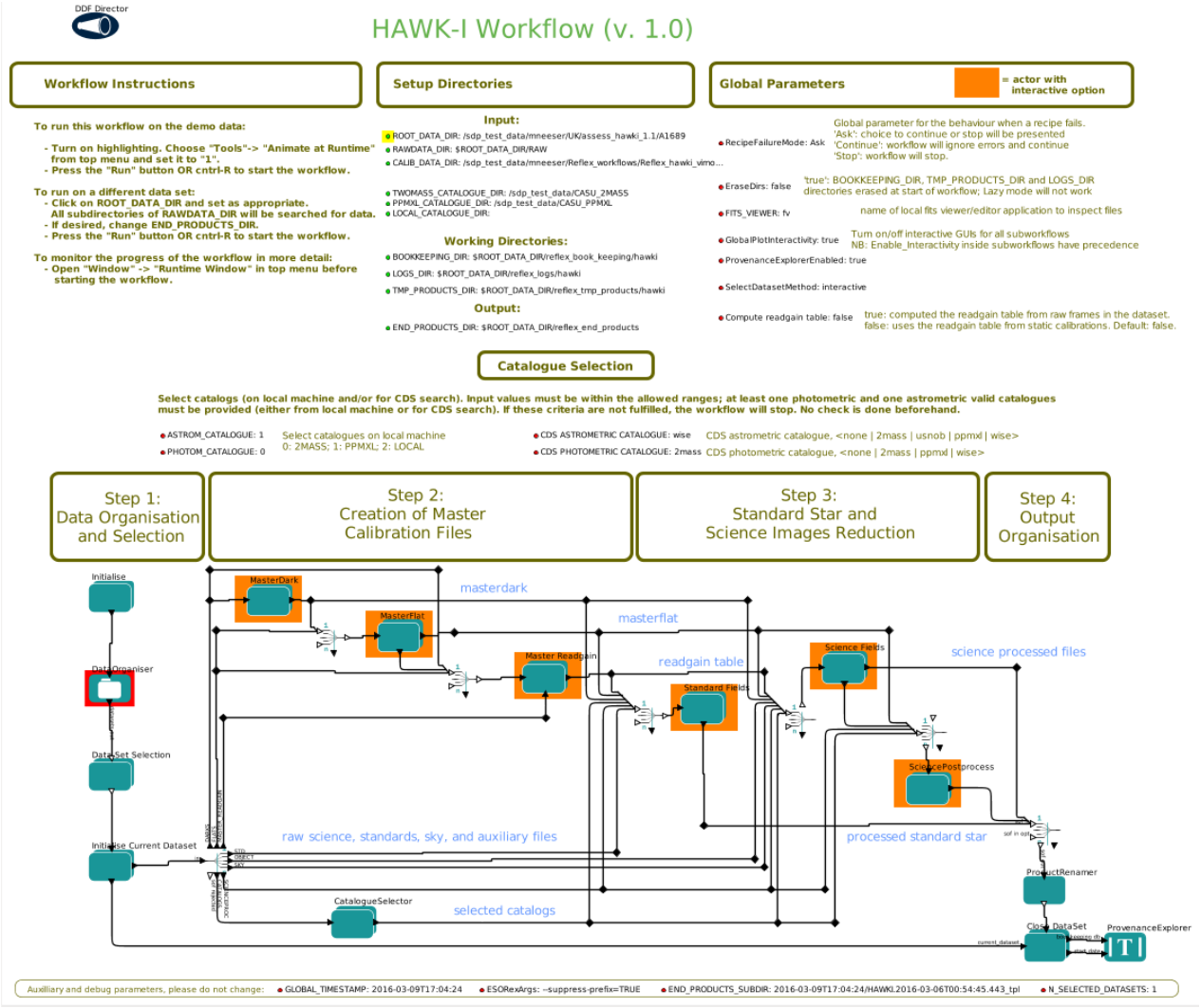

Figure 1. Figure captions are used to describe the figure and help the reader understand it's significance. The caption should be centered underneath the figure and set in 9-point font. It is preferable for figures and tables to be placed at the top or bottom of the page. LaTeX tends to adhere to this standard.

6.1 DataOrganiser AND DataSetSelection

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.

6.2 PROCESSING ACTORS

6.3 INTERACTIVITY

http://eso.org/pipelines http://eso.org/reflex

6.4 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.

7. STATUS

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.

8. SUMMARY AND CONCLUSIONS

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.

ACKNOWLEDGMENTS

This unnumbered section is used to identify those who have aided the authors in understanding or accomplishing the work presented and to acknowledge sources of funding.

REFERENCES

[1] Freudling, W., Romaniello, M., Bramich, D. M., Ballester, P., Forchi, V., García-Dabló, C. E., Moehler, S., and Neeser, M. J., "Automated data reduction workflows for astronomy. The ESO Reflex environment," *Astron. Astrophys* **559**, 423–424 (Nov. 2013).

[2] Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B., and Mock, S., "Kepler: an extensible system for design and execution of scientific workflows," in [*16th International Conference on Scientific and Statistical Database Management, 2004. Proceedings.*], *IEEE Conference Proceedings* (2004).