Mem. S.A.It. Vol. 80, 578

Memorie della

ESO Reflex: a graphical workflow engine for data reduction

Richard Hook[1], Marko Ullgrén[2], Martino Romaniello[1], Sami Maisala[2], Tero Oittinen[2], Otto Solin[3], Ville Savolainen[4], Pekka Järveläinen[4], Jani Tyynelä[4], Michèle Péron[1], Pascal Ballester[1], Armin Gabasch[1], Carlo Izzo[1]

[1] European Southern Observatory, Karl Schwarzschild Str-2, D-85748 Garching, Germany. e-mail: rhook@eso.org
[2] Observatory, University of Helsinki, Finland
[3] Space Systems Finland Ltd
[4] CSC–Scientific Computing Ltd, Espoo, Finland

Abstract. ESO Reflex is a prototype software tool that provides a novel approach to astronomical data reduction by integrating a modern graphical workflow system (Taverna) with existing legacy data reduction algorithms.
Most of the raw data produced by instruments at the ESO Very Large Telescope (VLT) in Chile are reduced using recipes. These are compiled C applications following an ESO standard and utilising routines provided by the Common Pipeline Library (CPL). Currently these are run in batch mode as part of the data flow system to generate the input to the ESO/VLT quality control process and are also exported for use offline. ESO Reflex can invoke CPL-based recipes in a flexible way through a general purpose graphical interface.
ESO Reflex is based on the Taverna system that was originally developed within the UK life-sciences community. Workflows have been created so far for three VLT/VLTI instruments, and the GUI allows the user to make changes to these or create workflows of their own. Python scripts or IDL procedures can be easily brought into workflows and a variety of visualisation and display options, including custom product inspection and validation steps, are available. Taverna is intended for use with web services and experiments using ESO Reflex to access Virtual Observatory web services have been successfully performed.
ESO Reflex is the main product developed by Sampo, a project led by ESO and conducted by a software development team from Finland as an in-kind contribution to joining ESO. The goal was to look into the needs of the ESO community in the area of data reduction environments and to create pilot software products that illustrate critical steps along the road to a new system. Sampo concluded early in 2008.
This contribution will describe ESO Reflex and show several examples of its use both locally and using Virtual Observatory remote web services. ESO Reflex is expected to be released to the community in early 2009.

1. The data reduction challenge

The European Southern Observatory (ESO) is currently operating a large suite of instruments

Send offprint requests to: R. N. Hook

covering the optical and the infrared, as well as the millimetre wavelength ranges. Although the responsibility for the quality of the scientific reduction of the data can only rest with the individual users, it is very difficult for users to be equally familiar with all the different observational techniques spanned by the ESO instruments at a level where general-purpose tools like IRAF and ESO-MIDAS can be effectively used. Instrument specific software, implementing carefully tuned algorithms, is therefore essential. Currently ESO aims to develop and export data reduction recipes for all VLT/VLTI instruments. These are based on the ESO Common Pipeline Library (CPL) and may be run offline using either the Gasgano graphical tool or the EsoRex command line tool. Recipes have the primary tasks of running as automatic pipelines within the dataflow system and being used to create products suitable for quality control (Silva & Peron 2004; Ballester P. et al. 2006).

The challenge at present is to allow the user greater flexibility to interact with the data reduction process and to study data products, both intermediate and final, in order to optimise the quality of the results. In addition it is desirable to reuse existing software as much as possible, both current pipelines and legacy software tools. The aim was to embed the ESO recipes within a flexible environment without the need to recreate a complete and expensive new software system. We believe that this approach has the potential to deliver a significant improvement for users whilst making optimal use of available resources.

2. Introducing ESO Reflex

The Sampo project, a three year effort led by ESO and conducted by a software development team from Finland as an in-kind contribution to joining ESO, has concentrated on developing a graphical user interface to run ESO data reduction recipes. The high level goals of the project were described in Hook et al. (2005).

The primary outcome of the Sampo project is an application called ESO Reflex (ESO REcipe FLexible EXecution workbench), in which the sequence of reduction steps is rendered and controlled as a graphical workflow. This approach allows users to follow and interact with the data reduction flow in an intuitive manner without the need for complex scripting. Figure 1 illustrates the look and feel of an ESO Reflex workflow. In this particular example, it is a reduction sequence to produce master calibrations for the VLT FORS2 instrument in MXU mode. The input files are at the top of the workflow (light-coloured boxes) and the data flow through the workflow to produce the final outputs at the bottom. The boxes in between the inputs and outputs represent the actual processors acting on the data, while the arrows mark the data flowing from one processor to the next.

ESO Reflex is based on a graphical workflow engine called Taverna that was originally developed for the eScience community in the context of the myGrid initiative in the United Kingdom (the project page is available at http://taverna.sourceforge.net). Once adopted after a survey of other available scientific workflow engines, Taverna was customised by the Sampo team to tailor it to the needs of astronomical workflows. These additions include a new interface for launching workflows, support for FITS files and interfaces to CPL, Python and IDL.

Workflows in ESO Reflex are easily edited and customised by simply adding or removing processors, the boxes in the middle of Figure 1, and connecting the appropriate input and output ports with arrows. The underlying workflow engine takes care of all the additional complexities linked to making the data flow through the reduction workflow as defined graphically by the user. The users of such a system are left to focus on their core task: making scientific sense of their data and exploiting them to the maximum.

The interface of ESO Reflex is not instrument specific and users are presented with the same look and feel independently of the actual instrument from which the data originated. ESO Reflex aims to provide most of the key elements for a scientific data reduction, including the following features:

- Convenient ways to select and organise data, based on code from the Gasgano ap-

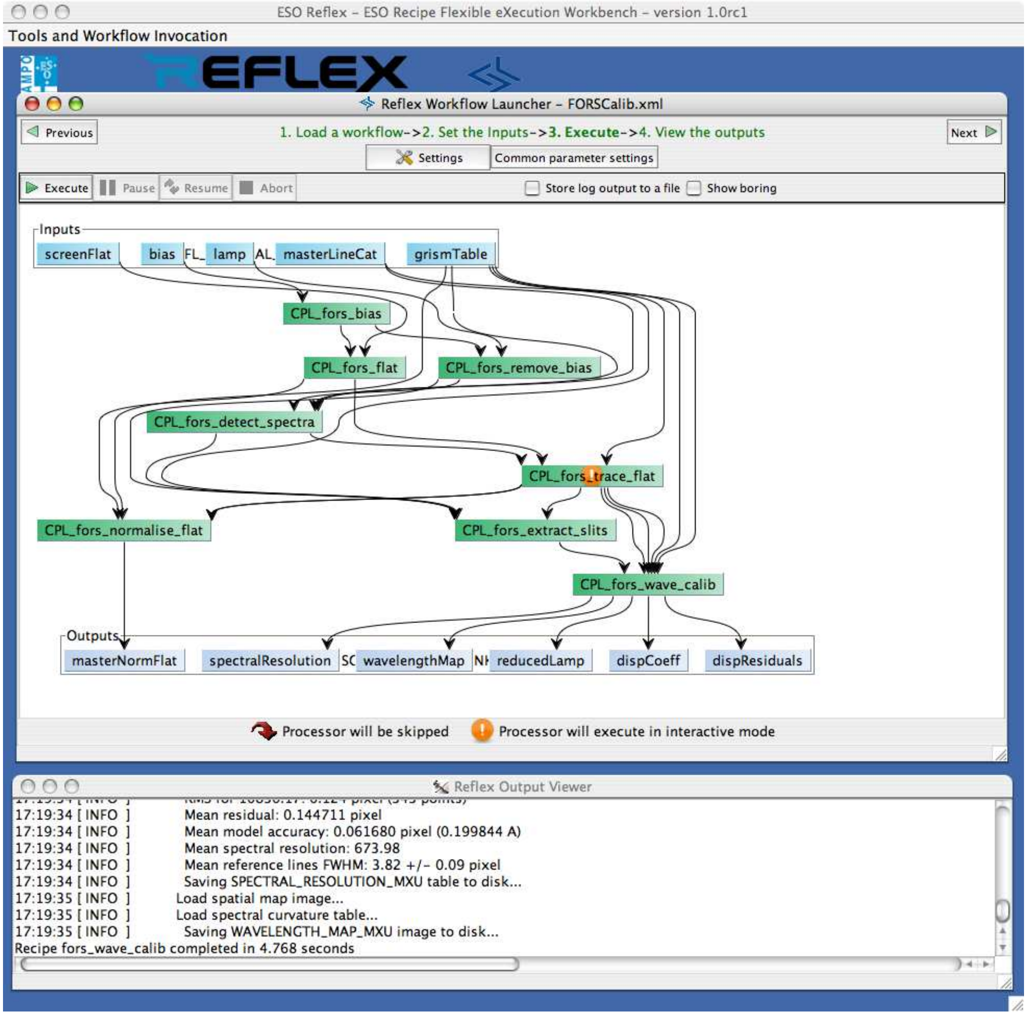

Fig. 1. Example of an ESO Reflex workflow based on calibration recipes for the VLT FORS2 instrument in MXU mode. The input data are represented by the light-coloured boxes at the top. The data flow through the processors in the middle section to produce the outputs shown at the bottom. The orange circular symbol indicates that one recipe will execute in interactive mode—this allows the user to inspect the input and output files from this stage of the processing and modify parameters if desired.

plication (http://www.eso.org/gasgano), to cope with the complexity of the headers of modern data.

– A CPL processor to include data reduction recipes for the vast majority of the data produced by ESO instruments into workflows. This dedicated processor is tailored to handling ESO data using CPL recipes and supports many extra features, including different processor modes (interactive, skipped etc.), as well as control of recipe-specific parameter values.
– Processors through which Python scripts, IDL procedures and shell commands can be included within workflows.

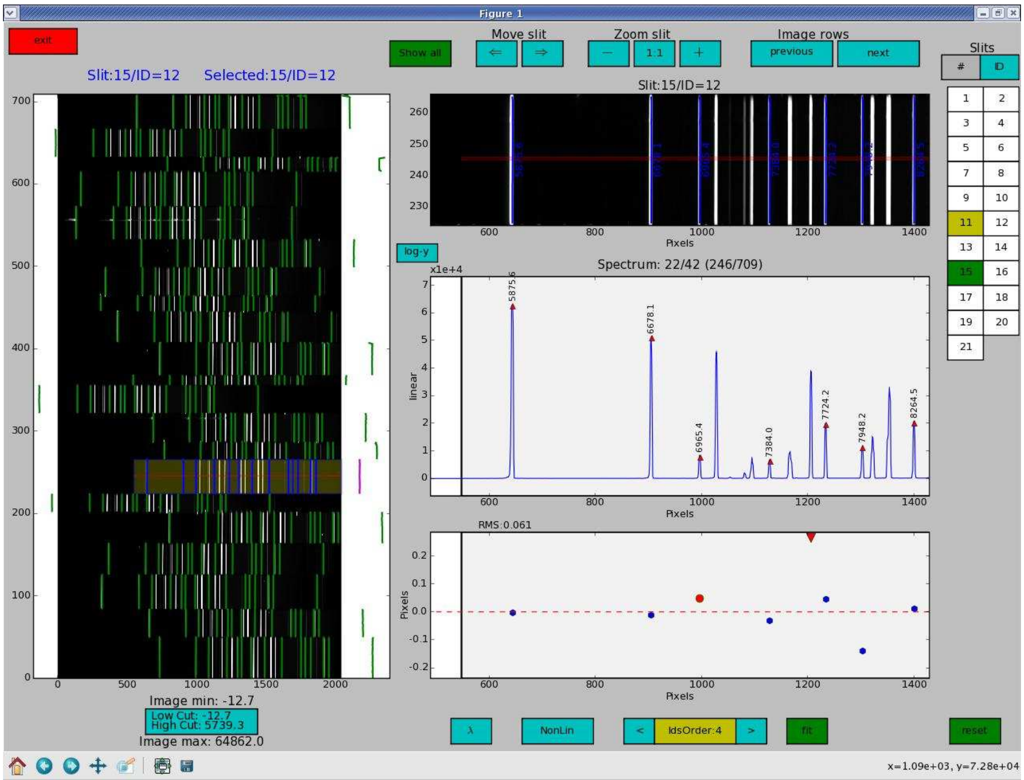

Fig. 2. An interactive Python-based tool to check and improve the wavelength solution of two-dimensional spectra. Calibration spectral lines, either from an arc lamp or from the night sky, are displayed slit by slit and can be included or excluded when computing the wavelength solution with a polynomial fit, the order of which can also be set interactively. This example shows an arc exposure from the VLT/FORS instrument in MXU mode.

– The possibility of basic flow control operators, such as conditional steps.
– Error handling: ESO Reflex catches errors returned by processors and offers options on how to proceed further, e.g. abort the workflow, reconfigure the offending processor and rerun it, proceed anyhow trying to execute the rest of the workflow.
– Skipping of processors and the possibility to allow optional steps.
– Automatic processing of lists of input files.
– Batch processing without the graphical user interface.
– The design of Taverna makes it very effective for building workflows that use web services such as those established within the Virtual Observatory. Experiments in this area have been successful and are described below, also see Järveläinen et al. (2008).
– A particularly important use of scripts is to analyse intermediate products within the reduction process. To illustrate this concept we have developed several interactive tools. A screenshot of such a tool, in this case to iteratively check and refine the wavelength solution of two-dimensional spectra is shown in Figure 2.

A graphical workflow system is, for many purposes, not as powerful as a well-crafted script. However, it is expected that the greater ease of use of a graphical workflow system will compensate for the loss of power, when com-

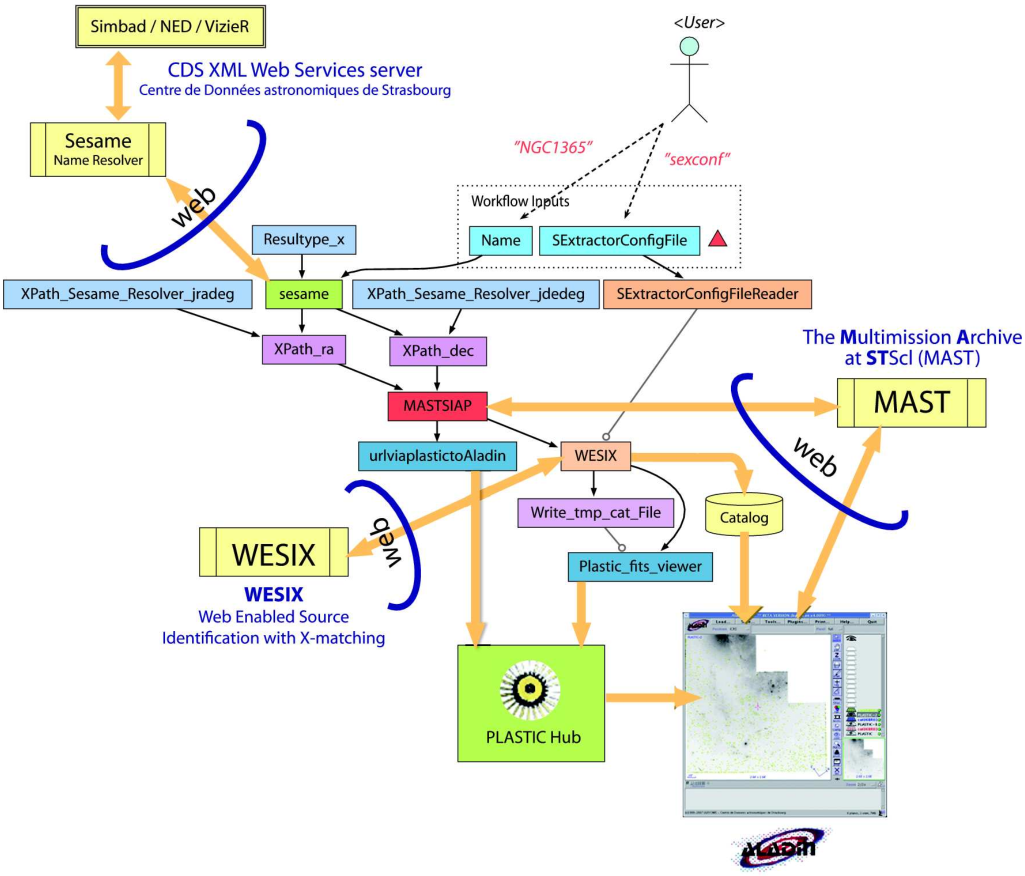

Fig. 3. Using ESO Reflex with astronomical web services. The workflow itself appears at the centre of the diagram and the three web services that are used appear at the edges. The final image and catalogues are visualised, via PLASTIC, using Aladin, and shown at the bottom.

pared to traditional scripting, for many users. Those unfamiliar with the reduction steps and options for a particular instrument and mode should find the clear graphical view, and simple interaction, provided by ESO Reflex to be very helpful.

3. ESO Reflex and Virtual Observatory services

ESO Reflex is based on Taverna and hence has excellent support for access to distributed resources and web services. In addition we have added PLASTIC support to allow convenient communication with client-side VO tools for visualisation. To test out these concepts a small project was conducted with the aim of making a workflow that carries out a realistic set of operations using existing VO services. The first step is to resolve an astronomical object name supplied by the user to a position on the sky using the Sesame service at the CDS in Strasbourg. The resulting sky positions are extracted from the text returned from Sesame and then used to query a remote SIAP data service — in this case to search for relevant Hubble data from MAST at STScI. The resulting VOTable is presented to the user through a graphical table viewer and the desired data sets are selected. These are then passed to a separate, remote, service (WESIX at the University of Pittsburgh) where the SExtractor tool (Bertin & Arnouts 1996) is run to catalogue the contents of the images. Finally the results, both images and catalogues, are returned to be locally displayed using a PLASTIC-enabled tool (Aladin) on the user desktop. Figure 3 shows this in schematic form.

4. Current status and future plans

At the time of writing, ESO Reflex is in a beta state and is expected to be released to the community at large in early 2009 along with appropriate workflows and tools. People interested in early access to ESO Reflex in conjunction with the instrument modes for which workflows have been developed, namely FORS spectroscopy and AMBER, should contact reflex@eso.org. Work is also in progress to enhance the data reduction recipes. The current algorithms are focused on processing calibrations and extracting the parameters required to monitor the health status of the instruments. While in some cases the resultant products are of adequate quality for immediate scientific analysis, this is generally not yet the case. To this end, the data reduction recipes are being made available in modular form to allow interaction with the intermediate products at scientifically meaningful points and to work seamlessly with ESO Reflex. The data reduction algorithms themselves are also continuously being extended with the long-term aim of allowing the creation of high quality science products on the user desktop.

References

Ballester P. et al. 2006, Proceedings of the SPIE 6270

Bertin E. and Arnouts S., A&AS, 117, 393

Hook R. N. et al. 2005, ESO Messenger, 120, 17

Järveläinen et al. 2008, in ASP Conf. Ser., ADASS XVII, ed. J. Lewis, R. Argyle, P. Bunclark, D. Evans, and E. Gonzalez-Solares (San Francisco: ASP)

Silva D. and Peron M. 2004, ESO Messenger, 118, 2