Astronomical Data Analysis Software and Systems XXIII
ASP Conference Series, Vol. 485
N. Manset and P. Forshay, eds.

Integrating Mass Scale Spectroscopic Processing

Nicholas A. Walton, Michael J. Irwin, Sergey Koposov, James R. Lewis, and Eduardo Gonzales-Solarez

Institute of Astronomy, University of Cambridge, Cambridge, CB3 0HA, UK.
Email: naw@ast.cam.ac.uk

Abstract. This paper briefly presents the challenges in implementing the data driven pipelines which will support the delivery of science data products from three major new spectroscopic surveys, namely the Gaia-ESO survey, the WEAVE multi-object spectrograph for the 4.2-m William Herschel Telescope, and the 4MOST multi-object spectrograph for the ESO VISTA telescope. We note the design solutions being implemented at the Cambridge Astronomical Survey Unit where an integrated approach in the delivery of a scalable data pipeline is being adopted. The design of the processing system is strongly science driven, which ensures that the analysis system delivers high quality data products to the science survey teams on the shortest possible time-scales, thereby allowing rapid scientific validation and exploitation of the data.

1. Introduction

The current decade heralds the development of several major new wide field multi-object spectroscopic (MOS) survey instruments, optimised to deliver spectroscopic observations of tens of millions of stars and galaxies. These instruments will facilitate progress over a huge range of science programmes particularly in the coming era of major facilities such as Gaia, eRosita and LOFAR.

The Cambridge Astronomical Survey Unit (CASU)[1], at the Institute of Astronomy, University of Cambridge, specialises in high throughput, science driven, data processing of major ground-based optical and infrared imaging surveys, including for WFCAM on UKIRT and the ESO VISTA NIR and VST optical surveys (see e.g. Irwin et al. 2004). Up to a TByte of data from these systems is routinely processed each night. The CASU pipelines scoped to support the generation of science products across a wide range of surveys, recent examples include generating the science data products for the Galactic Bulge (Saito et al. 2012), Magellanic Cloud (Cioni et al. 2011) and deep extragalactic (Jarvis et al. 2013) surveys.

2. Developing a Scalable Production Processing System, The Cambridge Approach

Building on this expertise in support of imaging surveys, CASU is developing processing and data management systems for a number of major new spectroscopic survey

[1]http://casu.ast.cam.ac.uk

instruments. In the following sections we describe the approach being taken to develop the processing and archive systems for these major new ground based multi-object spectrographs, noting how the implementation of the processing system to support a new survey utilising existing facilities provides a useful pilot system.

The data systems at CASU have been optimised to support the processing of optical and infrared imaging surveys. Many of the techniques used in these imaging pipelines are being adopted in the design of these next generation spectroscopic pipelines, for instance in the initial processing of the 2-dimensional data, in the overall handing of large data, and in the process used in developing, testing and documenting the analysis software. Here we note that CASU releases its analysis algorithms[2], future releases will include the full spectroscopic analysis suite.

2.1. CASU Data Systems for the Gaia-ESO-Survey

The Gaia-ESO survey (GES) (Gilmore et al. 2012)[3] is a 300 night survey of the Milky Way being undertaken on the ESO VLT utilising the FLAMES facility enabling observations with GIRAFFE and UVES. GES aims to observe some 100,000 stars in order to address a wide range of science topics related to the structure and formation of our Milky Way.

In order to support the analysis and distribution of the data both within the GES consortium and to the wider ESO science community, CASU have designed the data handling system, and are responsible for the operation of significant elements of this system.

CASU provides the processing architecture, the operational database, and the core Level-1 (L1) processing system. Higher level analysis (L2) is undertaken by partner expert groups who provide analysis for specific survey target types, for instance hot stars, cool stars, and so forth. The L1 and L2 science products are integrated into the CASU Gaia-ESO archive, initially for internal release, but later to the community via ESO as so-called ESO Phase 3 advanced science data products. Of note is the overall architecture and how this, combined with suitable interface control and a modular design, enables the analysis system to integrate successfully a range of expert-provided high-level analysis products. The internal data repository is available at `http://casu.ast.cam.ac.uk/gaiaeso`, this providing secure data access to the consortium. Figure 1 shows the observations overview available from the operational repository (here correct as of ~summer 2013).

2.2. WEAVE: A New Wide-field Multi-object Spectrograph for the William Herschel Telescope

WEAVE[4] (Dalton et al. 2012) is a major new multi-object spectrograph now starting construction for deployment on the 4.2-m William Herschel Telescope. With 1,000 fibres available over a π deg^2 field of view it will be a powerful tool to carry out a range

---

[2]The current CASUTOOLS release is at `http://casu.ast.cam.ac.uk/surveys-projects/software-release`

[3]see `http://www.gaia-eso.eu`

[4]see `http://www.ing.iac.es/weave`

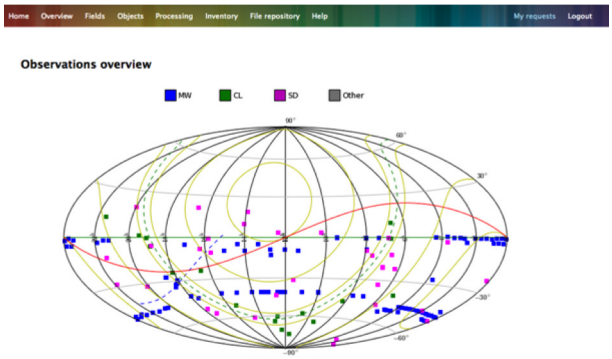

Figure 1. The CASU Gaia-ESO operational repository archive - showing here the status of observations as of ~summer 2013.

of galactic and extragalactic surveys. The WEAVE facility will begin on-sky surveys in 2017.

The processing architecture is based on that adopted for GES, where this will support the required up scaling to handle the significantly larger surveys from this instrument (10^7 objects in the WEAVE surveys c.f. with 10^5 objects from the Gaia-ESO survey). The processing methodology pays particular attention to issues such as sky subtraction, cross-talk suppression and minimal rebinning of spectra. Level 1 products include radial velocities/redshift estimate and first pass template matching to characterise objects prior to feeding these to the more distributed L2 systems for further more detailed analysis such as stellar abundance determination.

2.3. 4MOST: The 4-m Multi-Object Spectroscopy Telescope

4MOST[5] (de Jong et al. 2004) is an even more ambitious optical MOS to be deployed on the ESO VISTA telescope, beginning its five year survey programme in 2020. With 2,400 fibres over a ~4 deg^2 field of view, it will target some 30 million objects, stars in our Galaxy and deep extra galactic samples.

Figure 2 shows the architecture being adopted in the design of the 4MOST data flow system. This has many similarities with that for WEAVE. The L2 pipelines will be multiple, and will generated high level data parameters from the science 1-d spectra, specific to the object and survey type. For instance there will be a range of L2 treatment modules to derive accurate stellar abundances from the high resolution 4MOST data.

The systems being developed for GES, WEAVE and 4MOST all involve a multi-level processing flow with integration of, or interface to, (multiple) high level analysis (e.g. specific abundance analysis for cool stars) chains. In the development and operation the development approach takes care to manage all interfaces, control of processing meta-data, documentation and uses appropriate standards (e.g. use of FITs, keywords).

Full details of the evolution of the CASU spectroscopic processing developments are available at `http://casu.ast.cam.ac.uk`

[5] see `http://www.4most.eu`

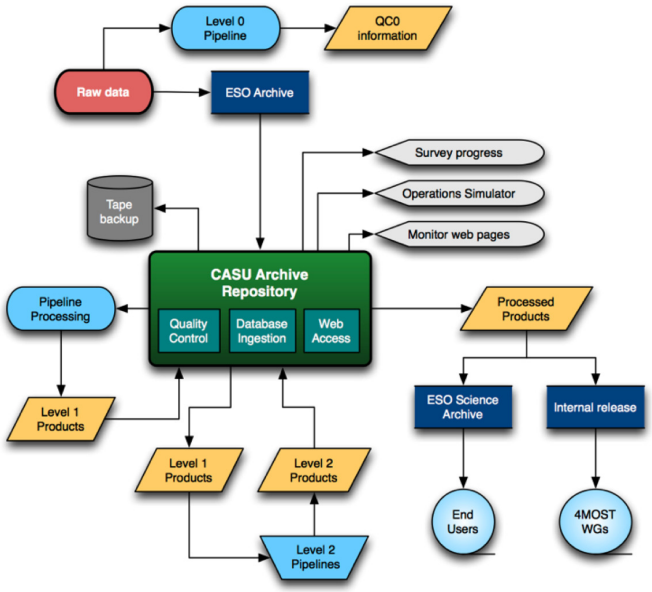

Figure 2. This shows the data flow that is being implemented for the 4MOST data management system. The core processing loop taking raw data from the telescope, running the 2-d pipelines to extract L1 one-d calibrated spectra, and the ensuing processing of those in the L2 pipelines is shown. The use of the operational repository is essential in managing the flow of intermediate products.

Acknowledgments. We would like to acknowledge the various partners in the WEAVE and 4MOST consortia who will be involved in the final development and operations of the science processing systems for WEAVE and 4MOST.

References

Cioni, M. L., et al. 2011, A&A, 527, 116
Dalton, G., et al. 2012, SPIE, 8446
de Jong, R. S., et al. 2004, SPIE, 8446
Gilmore, G., et al. 2012, The Messenger, 147, 25
Irwin, M., et al. 2004, SPIE, 5493, 411
Jarvis, M. J., et al. 2013, MNRAS, 428, 1281
Saito, M., et al. 2012, A&A, 537, 107