*Astronomical Data Analysis Software and Systems XIII*
*ASP Conference Series, Vol. 314, 2004*
*F. Ochsenbein, M. Allen, and D. Egret, eds.*

# Extending ORAC-DR for Offline Processing of ESO, INGRID, and Classic Cam data

Malcolm J. Currie

*Starlink, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, U.K.*

**Abstract.** ORAC-DR—a flexible reduction pipeline—was originally developed by the Joint Astronomy Centre for real-time inspection of reduced data at its telescopes. Starlink is extending ORAC-DR to process at home institutions data from other observatories, notably ESO, whose instruments make no provision for ORAC-DR. I outline the problems encountered and solutions implemented or proposed to apply ORAC-DR to the infra-red instruments ISAAC, NACO, INGRID, and Classic Cam.

## 1. Introduction

ORAC-DR is an intelligent, data-driven, tailorable pipeline system, developed at the Joint Astronomy Centre (JAC). Object-oriented Perl data-reduction recipes use Starlink applications to process the bulk data, invoked within 'primitives', each of which correspond to an astronomically meaningful step in the reduction. It has general code for common operations such as managing calibrations, flexible data display, and data-format conversions.

Before the work described here, ORAC-DR already supported a wide variety of instruments and techniques at UKIRT, JCMT and AAO (Cavanagh et al. 2002 and references therein). At these institutions the data collection cooperates with ORAC-DR, so the sequence of integrations match the reduction recipes, and ORAC-DR-specific metadata are present.

Starlink realised that ORAC-DR had the potential to reduce data from telescopes even without this symbiosis, or face-to-face dialogues with instrument scientists, available in an observatory setting. This is particularly important for UK astronomers recently exposed to complex ESO instruments. It has potential for Virtual Observatory (VO) applications, where an archive stores the raw data and the astronomer wants a reduced mosaic or fully calibrated spectrum. The reduction recipes can encapsulate knowledge of the instrument signature and reduction techniques, freeing the astronomer to concentrate on the science.

This paper describes some of the problems and solutions adopted to apply ORAC-DR to some infra-red instruments ISAAC and NACO at ESO; INGRID at the ING, La Palma; and Classic Cam on Magellan.

## 2. Problems and Solutions

### 2.1. Metadata

ORAC-DR is driven by FITS metadata. These comprise steering headers like the observation and group numbers, and recipe name; and recipe-specific metadata, such as filter, exposure time, and grating name. ORAC-DR translates the latter type of header, sometimes in combination, into common internal headers with standard meanings and units, thereby insulating the recipes from instrument-specific headers.

While the observation number is usually present, the other steering headers are not. As ORAC-DR works sequentially, preprocessor C-shell scripts determine from a combination of headers or assumptions based upon web-based information, the delineation of the groups (each group being a related set of observations). For example, the `HIERARCH.ESO.TPL.EXPNO = 1` in ESO data implies the start of a group; as does when any of the main attributes like filter, and exposure time change in Classic Cam headers. The scripts insert the few required steering keywords into the metadata.

There is a close mapping of ESO templates to existing ORAC-DR recipes. However, it's not always one-to-one, for example it's not clear whether to self-flat or use a separate flat for a dithered infra-red sequence. These are different recipes, but the same template. The user can always substitute the appropriate recipe on the ORAC-DR command line, or edit the `RECIPE` header. Commonality between dictionaries at ESO, does permit general infrastructure, subclassing where necessary to specific instruments. Code reuse is an ORAC-DR mantra.

The quality of FITS headers is extremely variable, and many cases omit vital data for recipes, let alone provide an accurate record of the instrument status and something at all adequate for VO use. The Classic Cam headers were particularly spartan. Even the detailed ESO headers had some metadata needed by the recipes missing, such as the grating dispersion. Appropriate values were usually found in or derived from web pages and manuals, which for ISAAC were very good. If only other observatories were as diligent. However, occasionally I had to consult an instrument scientist, more so for Classic Cam and INGRID. Dialogue also ensued whenever there were two headers ostensibly presenting the same attribute, but with different value, or the meaning of a header was unclear or its units were not specified.

Headers change as experience with an instrument grows and a pipeline must cope. Where available, data dictionaries help greatly, but their history is not readily available to the external developer. It would help if the CVS repositories of the dictionaries were made publically accessible. For ESO data I extended the translations empirically, by testing with data over a wide range of observation dates. Judging by the experience at UKIRT, no doubt the existing translations are not comprehensive and will require more exposure to users' data over an extended period. Some expected headers were not always present, so the translations must be flexible and search a few headers in turn. The data dictionaries don't include the dependencies.

The ESO headers include hierarchical keywords, which required modification of the Perl package `Astro::FITS::Header`. Starlink applications had already handled these headers for many years.

### 2.2. File Naming

ORAC-DR file naming mostly follows the JAC conventions, although there is some provision for others. ESO use the UT date and time in one of its naming conventions, while ORAC-DR currently needs the observation number. Thus the preprocessor scripts rename the files to have JAC-like nomenclature, while allowing for times spanning midnight UT. A goal is for ORAC-DR to accept such names, and to allow selection of observations made between certain times.

### 2.3. Test Data

Some of the ports were instigated by users with data to reduce. Data from an individual user is highly selective. There may be no clues as to how representative they are. The INGRID, Classic Cam, and NACO pipelines therefore only have limited testing. NACO does, however, benefit from the ISAAC pipeline.

Accessing data from the ESO archive is not straightforward when you do not know which object you want, and whether the data you select is representative or pretty. This demands closer co-operation with the instrument scientists and archivists. It would be useful to collate representative test data for regression testing of ORAC-DR.

### 2.4. Instrumental Idiosyncrasies

Every detector and instrument has its own properties. For example, ISAAC has spatial distortion, electronic ghosting, variable bias. NACO—at least in the data seen thus far—has large swathes of very noisy pixels. INGRID has a multi-extension FITS file containing pre-exposure and post-exposure images. Each instrument has a different non-linearity correction. Polarimetry masks and dither patterns vary between instruments.

It is easy to insert new steps into a recipe or have instrument-specific steps without affecting existing primitives or other instruments' reduction. It's then a question of finding a suitable algorithm. All were solved using existing atomic Starlink applications. While the solutions were not especially sophisticated, they seem fine judging by the photometric accuracy and quality achieved.

## 3. Recipe Development

The processing demands of the instruments new to ORAC-DR required new recipe code, which in some cases benefited the already supported instruments. Nevertheless much primitive code was reused. The highlights of these are listed below.

- Automatic wavelength calibration of non-thermal infra-red ISAAC spectra.
- Support for dithered target and offset spectral beams.
- New recipes for calibration-lamp flats.
- Combine multiple dark and arc integrations, and process multi-filter groups of flat-field frames.
- Correction for spatial distortion, variable bias, and electronic ghosting.
- Cope with several more observation sequences, such as various object-sky patterns for chop-sky imaging, not just *SOSOS*... and uses time-based interpolation for its modal sky subtraction.

## 4. Conclusions

- It is viable to generate an ORAC-DR pipeline producing publication or near-publication quality results for non-ORAC observatories. Most recipes were used 'off the shelf'.
- It is possible to achieve this in days or weeks rather than months or years in the simpler cases. The exact time depends on the number of new recipes needed, the quality of the metadata, the instrument complexity and special characteristics. To fully support an instrument like ISAAC in all its modes is still about six months' work, but extending to other ESO infra-red instruments will be much less. As the recipes encompass more instruments there will be fewer new problems to tackle.
- Comprehensive testing is not possible for the external programmer. Fine tuning for secular changes, and rare or pathological cases relies on user feedback.
- Standards for metadata need to improve markedly at certain institutions. This includes providing clear comments and units. A data dictionary is desirable too, preferably with interdependencies included. The dilettante culture must change for the VO too. Time to name and shame!
- It is better for the metadata to be too verbose, and record more instrumental and sequence data than is sufficient. Such information can also be useful diagnostics to secular or sudden changes of instrument properties. The pipeline can only take account of these, if it can recognise them. If observatories could be persuaded to add a few additional steering headers, the pre-processing step could be avoided. ORAC-DR offers a cheap pipelining system for observatories without resources to create their own.
- Astronomers do not like black boxes, and will surely wish to tinker with the provided recipes. While ORAC-DR offers some tools for this, it needs a configuration system and command-line overrides for recipe variables to permit easier tailoring of recipes by astronomers. This is in development. The ORAC-DR programmers' manual (Jenness & Economou 2001) needs updating and made astronomer friendly.
- It would be useful to have collated infra-red standard-star data available from data centres, accessed by cgi scripts, as already happens for spectroscopy calibrations in ORAC-DR, or via a web service.

**Acknowledgments.** Starlink is funded by the Particle Physics and Astronomy Research Council (PPARC) and managed by the Space Science and Technology Department (SSTD) of the Central Laboratory of the Research Councils (CLRC).

## References

Cavanagh, B., Hirst, P., Jenness, T., Economou, F., Currie, M. J., Todd, S., & Ryder, S. D. 2003 in ASP Conf. Ser., Vol. 295, ADASS XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 237

Jenness, T. & Economou, F. 2001, Starlink User Note 233, Starlink Project, CCLRC