Astronomical Data Analysis Software and Systems VI
ASP Conference Series, Vol. 125, 1997
Gareth Hunt and H. E. Payne, eds.

The SRON-HeaD Data Analysis System

C. P. de Vries

Space Research Organization Netherlands (SRON), Sorbonnelaan 2, 3584 CA Utrecht, Netherlands, E-mail: C.deVries@sron.ruu.nl

Abstract. A data analysis system has been developed at SRON, which has been designed to allow rigorous control of the quality of its processed data products. In order to fulfill this requirement, all data processing steps are recorded in a central database. The system will initially be used for analysis of *SAX*-WFC and *XMM*-RGS data at SRON.

1. Introduction

A critical requirement for data analysis systems used for massive routine data processing is that the system is able to deliver processed data of controlled quality in an automatic fashion. In order to control quality of data, it is necessary to have the ability to trace the heritage of all data products. This means recording all parameters of all steps which lead to the establishment of the final data products (Figure 1). To check processing status and initiate subsequent processing steps, one should easily be able to generate overviews of all available data and intermediary products based on data descriptions and processing heritage.

The SRON-HeaD (SRON High Energy Astrophysics department Data analysis) system has been developed to fulfill these requirements, and is based on earlier experiences with the *CGRO*-Comptel data analysis system (de Vries 1995).

2. Requirements

The following basic requirements were defined:

- Full traceability of data processing. Storage of all parameters of all processing steps, including complete software configuration.
- Complete catalogue of all available data products. Proper user-interface for manual processing and access to data descriptions and data heritage.
- Automatic processing based on data catalogue and processing status.
- Automatic archiving/retrieval of bulk data from mass storage devices.
- Use FITS format data files, where possible.
- Allow external analysis packages (e.g., FTOOLS, IDL, etc.) in the system.
- Separate "test environment" for testing of all system aspects and data processing programs. Capable of running on a variety of UNIX systems.

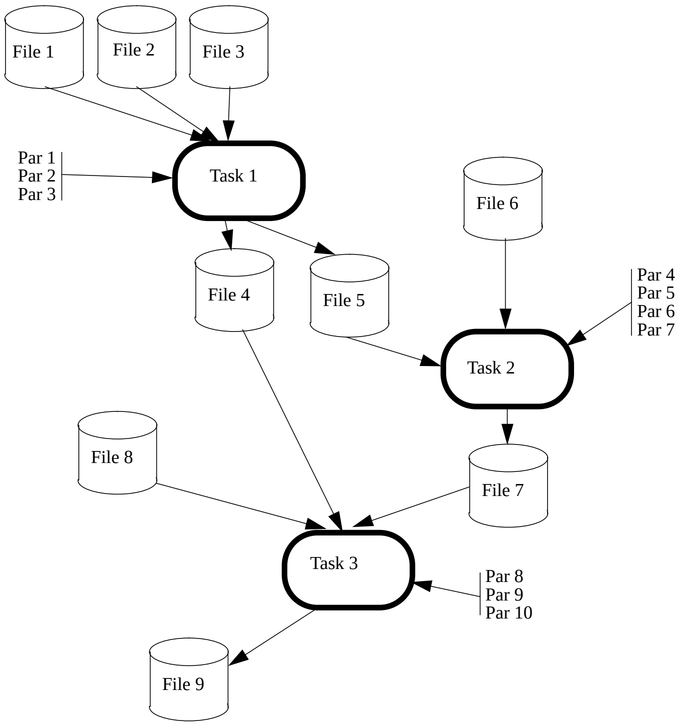

Figure 1. A processing pipeline is a sequence of tasks with several input/output files and parameters, which pass data from one task to the next. Heritage of any file (e.g., file 9) can be established by recording all input/output files and parameters of all steps and by uniquely identifying each task, requiring thorough configuration control.

3. Implementation

The core of the system is the recording of data descriptions and data heritage in the database, where this information may be queried via user interfaces or the routine processing pipeline to start new jobs (Figure 2). The dataset heritage consists of actual processing parameter values and the software configuration used. Since data processing parameters are available either in the FITS headers or from input parameter files used by the generating programs, these parameters can be recorded after actual data files have been created. This means that no connection to a database is required during data processing and that external packages can easily be incorporated into the system. In addition, externally generated (FITS) data files can easily be imported into the system, as well.

The basic processing module is a "task" or program executable, called from a script or "job," which may also call other tasks. The script defines the control flow and communication between tasks within a job. The job script may prompt the user for task parameter input. Actual parameters are passed to tasks via IRAF format parameter files. Automatic recording of processing parameters is done for each output file at the end of job processing.

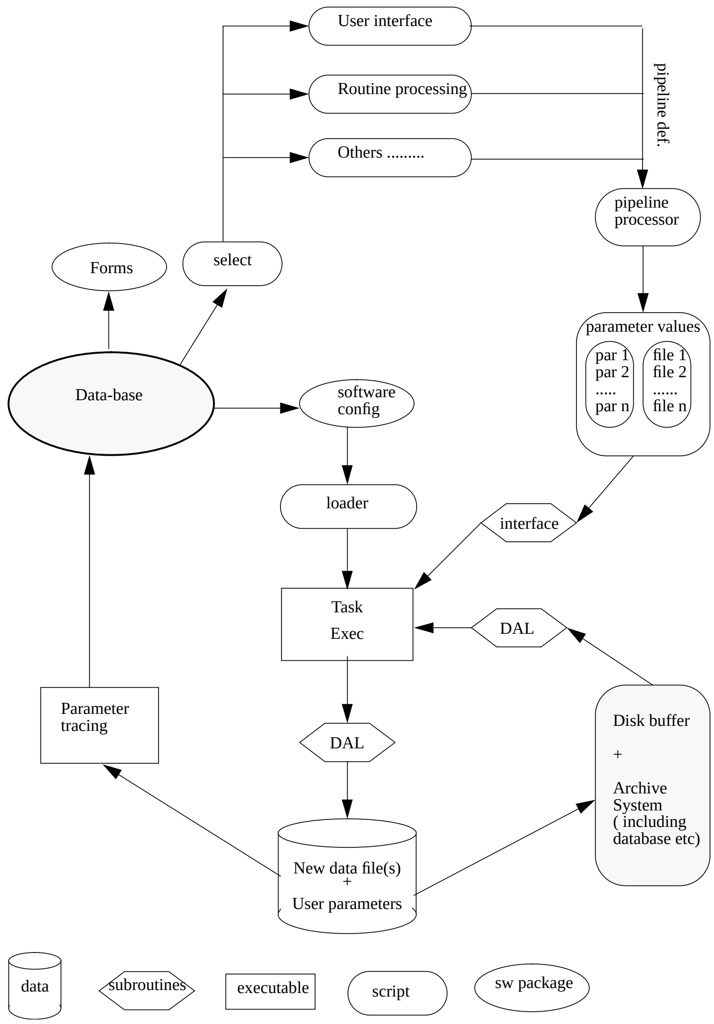

Figure 2. Top level functional breakdown of the HeaD system.

An Oracle client-server database architecture serves as the central database system (RDBMS), which holds the data catalogue and heritage, processing parameters description and software configuration. Software configuration at the system level is maintained through use of the RCS system, and use of this system is enforced by the appropriate user interfaces.

The Tcl/Tk package plus extensions (TclX,OraTcl) is used to create the windows-based user interface (Figure 3). On-line help is available by means of a Tcl/Tk HTML viewer. Direct selection of data files for input to the appropriate user interfaces can be made by selecting from lists resulting from the execution of SQL procedures. These can be taken from a library of procedures which base selections on a variety of data descriptions, heritage, or processing status. In addition, Oracle-Forms can be used for direct database queries.

Data processing "jobs" can be initiated manually by explicitly entering task parameters for specific tasks, or automatically through manual or automatic

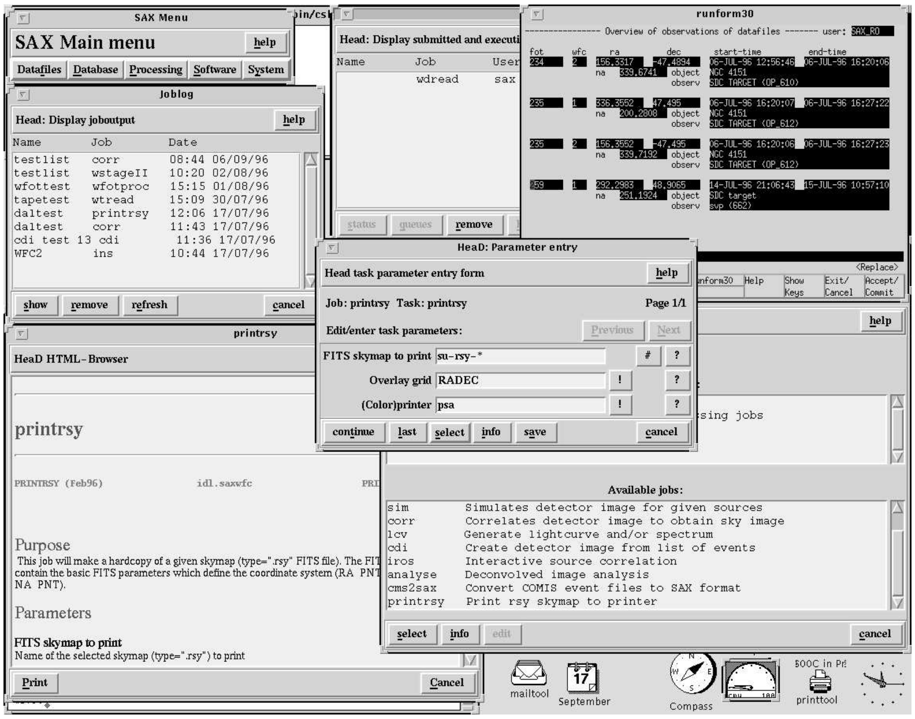

Figure 3. HeaD user interface.

marking of individual datasets or dataset types for further processing. In that case, special database tables define the processing flow.

A data access layer (DAL) is available which separates the actual scientific code from basic data I/O, allowing for greater system portability. This layer is partially composed of the FITSIO library, modified to allow for communication with the archive system, and specially developed routines.

Currently the system contains specially developed data processing programs as well as tasks taken from general packages like FTOOLS, IDL, SAOimage, etc. The system has been initially implemented on Sun (SunOs, Solaris) and HP (HP-UX) systems.

More information can be found on HEAD Home Page.[1]

References

de Vries, C. P. 1994, in ASP Conf. Ser., Vol. 61, Astronomical Data Analysis Software and Systems III, ed. D. R. Crabtree, R. J. Hanisch & J. Barnes (San Francisco: ASP), 399

[1] http://ws13.sron.ruu.nl:8080/head/Welcome.html