Starlink Software

David Giaretta, Malcolm Currie, Stephen Rankin

Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK

Peter Draper

Durham University, UK

Norman Gray

University of Glasgow, UK

David Berry

University of Central Lancs, UK

Mark Taylor

University of Bristol, UK

Abstract. We demonstrate the latest Starlink software, including ORAC-DR pipelines, new Java applications, and distributed pipeline processing.

1. Data Model

Starlink (`http://www.starlink.ac.uk`) has been developing new infrastructure and applications in Java, to supplement its large collection of Fortran and C applications in order to fit in with the developments in the Virtual Observatory. This infrastructure has been based on our best guess for the VO data model. As this real VO Data Model is developed we expect to modify the infrastructure to support it. Figure 1 shows how the components fit together.

Other VO standards will be supported as they are developed, for example the VOTable standard. Starlink provides what is almost certainly, at the time of writing, the most complete support for VOTable. An example of its use is the TOPCAT application (`http://www.starlink.ac.uk/topcat`), shown in part in Figure 2.

2. Infrastructure

2.1. Tables

TOPCAT is an application which uses the powerful TABLE interface which we have developed. This interface is designed to allow us to write applications which

832

HDX—provides way to handle hierarchical structured distributed data, using XML, FITS, HDS and other formats, adding metadata

Section of a proposed International Virtual Observatory Alliance Data Model

NDX—handles
- arbitrarily sized
- distributed
- N-Dimensional data
- *automatically tracking data errors and data quality*

Table—uniform access to distributed tabular data RDBMS tables (mySQL, Postgres, Ingres, Oracle etc), FITS, VOTable (full functionality) etc

WCS—state-of-the-art astrometry library, able to combine data in different coordinate systems and units. TimeFrame support is being added
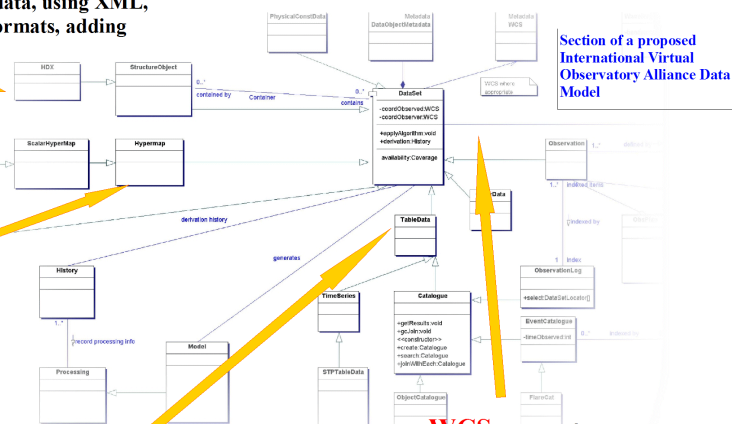
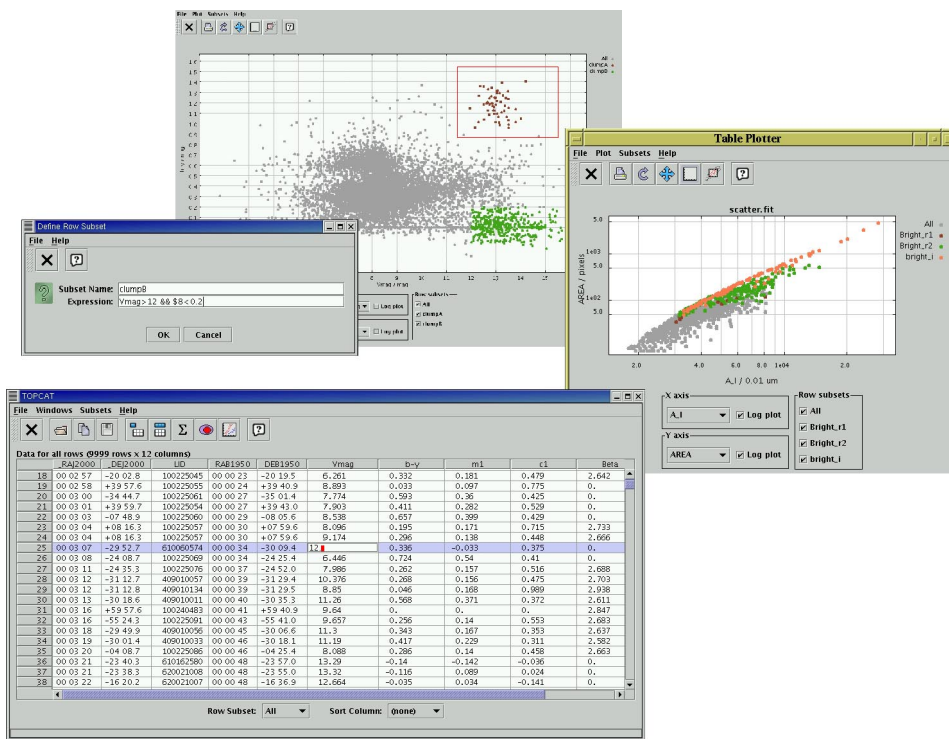Figure 1.   Relationship of Data Model to infrastructure
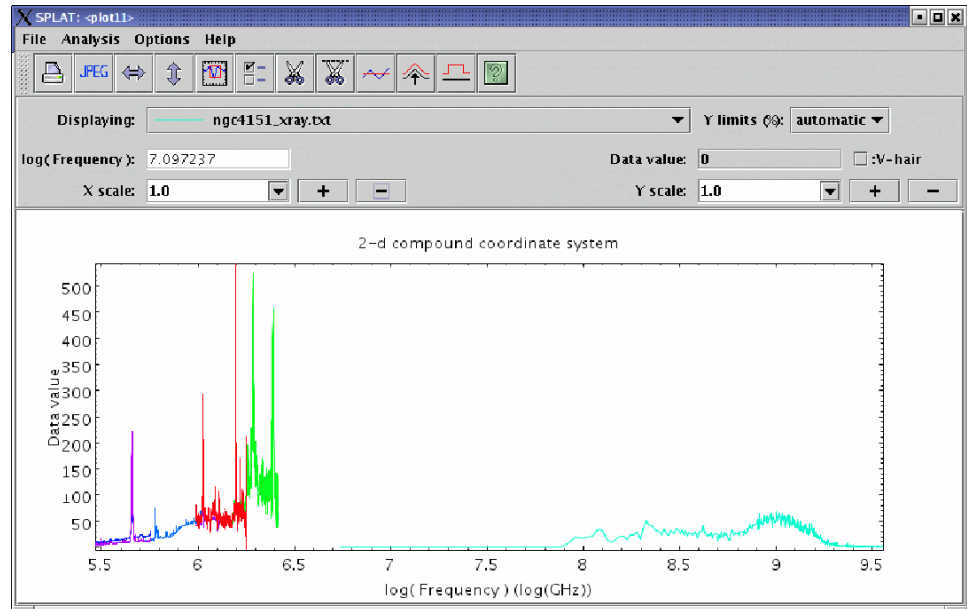
Figure 2.   TOPCAT

Figure 3.   Plot produced by SPLAT by dragging and dropping spectra from a variety of sources with axis coordinates including mm, Angstroms and KeV - AST automatically matches axes

can use just about any underlying format, as long as it is conceptually tabular. Such formats range from FITS tables, ASCII as well as BINARY, VOTables, RDBMS systems such as ORACLE or PostGRES, and it is fairly easy to add support for new formats.

2.2.   NDX - N-dimensional data including images

NDX deals with images and data cubes and higher dimensional data arrays. The important aspect here is that, as with TABLES, a variety of underlying formats can be used - in this case data which is conceptually array-like. Unique aspects of NDX include (1) the system is designed to deal with arbitrarily large datasets (2) the system automatically keeps track of data errors and data quality, if they are present. Both (1) and (2) are increasingly being recognised as vital for the Virtual Observatory era.

2.3.   Generalised World Coordinate Systems

Starlink has developed AST (`http://www.starlink.ac.uk/ast`) to deal with the plethora of World Coordinate Systems and their many encodings, and automatically mapping between them. An example of the capabilities this enables is demonstrated by dragging and dropping images onto each other - automatic coordinate mappings and data resampling aligns the data, as shown in Figure 3.

2.4. HDX - Structured data

Inter-related images, tables, text, metadata will become increasingly important.HDX is designed to provide a general purpose container. It includes the possibility of attaching updated metadata, including for example updated astrometry, to read-only image data. Nevertheless HDX has been designed as a fairly minimal container, easily extendible to include more sophisticated things, such as RDF, as they are adopted to capture increasingly complex interrelationships.

3. Classic Applications and ORAC-DR Pipelines

In addition to the new Java applications we have a large collection of Classic applications which provide a wide range of capabilities. These applications supply the processing engines used by ORAC-DR. ORAC-DR recipes capture the expertise needed to reduce instrumental data properly.

We are developing recipes for ORAC-DR to allow users to process data from a number of ESO instruments, where there is currently no other pipeline available to users, in particular archive users. We believe that we are showing good, consistent results with our recipes. Of course the final result can also provide estimates of errors and data quality, pixel by pixel, because the Starlink applications, Classic as well as the new ones, automatically keep track of these.

4. Distributed Processing and mini-GRIDs

We use a variety of protocols to allow us to distribute processing. These include Web Services, RMI and Jini. The latter has been used as the basis of a way of using surplus processing power in a min-GRID. PC-type machines on the network can be booted with our bootable CD which contains a Linux installation with all Starlink applications and appropriate JavaSpace configuration - nothing need be installed to the local hard disk. Client machines can submit tasks to run either serially or in parallel to the JavaSpace. These are automatically performed on the various servers that have been set up.