STATUS OF THE NEW ARCHIVE INFRASTRUCTURE FOR SPECTRA AT ESO

Markus Dolensky

The European Organization for Astronomical Research in the Southern Hemisphere

ABSTRACT

The community expects seamless and uniform search capabilities across missions and archives. This is a report on ESO's effort to approach this goal for various types of spectra in its vault and its links with the International VO Alliance (IVOA) and EURO-VO. It touches on aspects which are of interest to those wishing to publish spectroscopy data as well as to potential consumers and other archive facilities.

Key words: Astronomical data bases: miscellaneous; Technique: spectroscopic.

1. GETTING SCIENCE READY DATA FIRST

ESO has got numerous instruments in operations which do generate spectroscopy data on a nightly basis. Some are pure spectrographs producing sophisticated 3D data incl. integral field unit observations or Echelle spectra. Some are imagers supporting also certain spectroscopy modes, see fig. 1.

Just putting the raw frames and engineering telemetry into an archive is of limited usefulness for archival research purposes. Hence, reduced spectra produced by two pipelines are currently ingested, namely those of the High Accuracy Radial velocity Planet Searcher (HARPS), fig. 2, mounted on the 3.6m telescope on La Silla and UVES, fig. 3, an Echelle spectrograph mounted on unit telescope Kueyen on Paranal.

There are various further sources of science ready data such as:

- ESO Large Observing Programs
- Surveys: zCOSMOS, GOODS, K20 etc.
- ESO Public Surveys: VST, VISTA

Starting from observing period 75, PIs of Large Programs at ESO are requested to deliver to ESO final data products

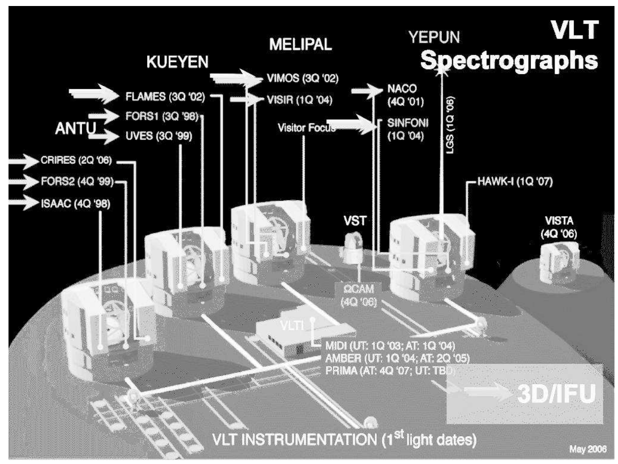

Figure 1. ESO Spectrographs and instruments with spectroscopy capabilities at the Very Large Telescope (VLT) on Paranal, Chile.

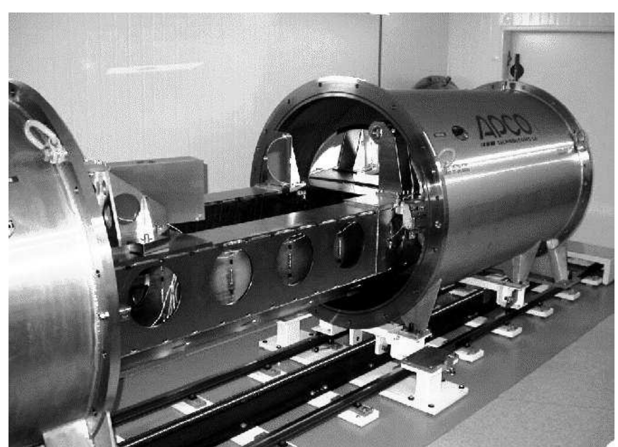

Figure 2. High Accuracy Radial velocity Planet Searcher (HARPS) pipeline products are publicly available through archive.eso.org since Q2/2007.

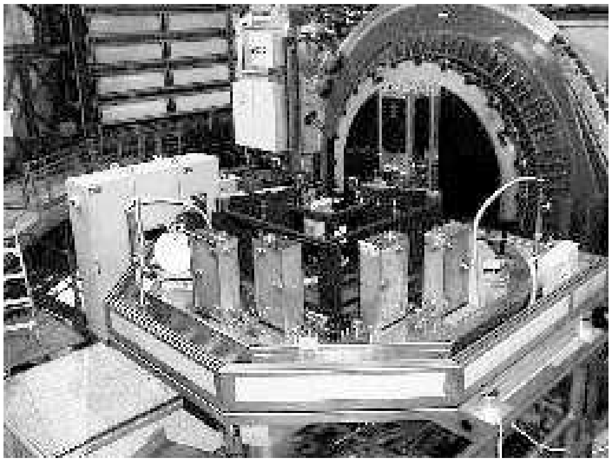

Figure 3. UVES Instrument: Its high-resolution, pipeline reduced spectra, (res. 110.000, 300-1100 nm) will become publicly available in Q3/2007.

at the time of publication of their results. Such advanced data products (ADP) from upcoming Public Surveys with VST and VISTA will also be ingested and distributed. To this end, ESO is developing a web-based product submission and upload interface becoming available in 2007.

The motivation for data providers and the community at large is manifold:

- enhanced legacy value of ESO data, facilitating further scientific exploitation
- uniform distribution of data products through the ESO Science Archive facility
- high visibility of data products and scientific results
- global broadcast to VO resource registries
- archive publications rate is about 20 times higher than for raw data

Reduced and therefore science ready data need to be archived and characterised. The ESO Science Archive Facility started ingesting such data in the 2nd quarter of 2007. Characterising observations generally means to store certain parameters in a data warehouse for flexible and swift query access.

The kind of meta data obviously depends on desired functionality of the query interface. A survey among archive users (Delmotte , 2006) helped to pin down typical needs and current shortcomings such as ...

- search by list of objects
- visualisation and previewing capabilities
- association of calibration frames
- VO compliance: for SED building; code re-use
- search by object category

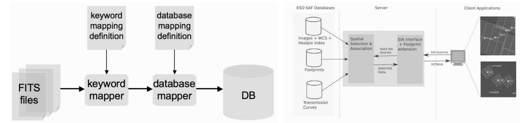

Figure 4. Left: Data homogenisation and characterisation. ; Right: A typical query scenario involves a series of transactions across the network. Above example combines a number of pre-populated databases (left), utilises an archive independent transfer protocol and data format (middle) and enables rich graphical representation (right)

2. DATA HOMOGENISATION

Moving from a technical observation log to an archive warehouse which characterises observations in an instrument and observatory independent way is a demanding process. It requires a machinery fig. 4 that homogenises heterogeneous data sets from varying origin and at different reduction levels into an integrated search engine.

Special software tools had to be implemented to help with the internal submission process of science ready data to the archive (Leoni , 2006), (Slijkhuis , 2006). At the core is a meta data extraction utility called `MEx` which is built with support from the European Commission funded VOTech project. As shown in fig. 4 the homogenisation starts from data usually stored as FITS files. The data organisation typically follows a more or less implicit native data model which depends on the observation mode and analysis code. The key point is that an archive usually has no control over its input. The homogenisation happens when mapping a specific set of key parameters to a defined data model. This entails computing and deriving values and applying physical unit conversions. The result of this step is made persistent in a database system. At this level observatory specific policies (business logic) and bookkeeping procedures are applied. The resulting database structure serves as the basis on which standard access protocols such as the Simple Spectral Access (SSA, 2007) protocol can be based upon.

Obviously, the details of this machinery are hidden from archive users and therefore one will note its existence only when it fails to resolve queries and to render results as expected.

3. PUTTING IT TOGETHER

Fig. 4 shows a typical query scenario and the underlying components. Efficient positional search is based on a spatial indexing algorithm called Hierarchical Equal Area isoLatitude Pixelization (Healpix). It partitions the celestial sphere to pre-selected small subsets which dramatically speeds up browsing among millions of frames. Associated are World Coordinate System (WCS) parameters for each exposure which are extracted as part of the characterisation described above. In the given query context the WCS is used to compute instrumental footprint overlays (Bonnarel , 2007). Meta information as well as actual data frames are returned using a widely supported protocol like (SIA , 2004) and data format (Ochsenbein et al. , 2004) facilitating the federation of cross mission query results.

From an archive research perspective the evolution of the interface started from more or less independent instrument- and project-specific query forms. There is now an integrated search page (DelmotteII , 2006) across instruments and telescopes and the next step is to integrate science ready data from projects and surveys.

4. TOWARDS AN SED BUILDER

The ultimate archival research tool is a system providing means to assemble the Spectral Energy Distribution (SED) of a given object. An initial step towards such an SED builder is the new magnitude to flux `Mag2Flux` conversion web service fig. 5 [1]. It dynamically accesses instrument transmission curves from the ESO exposure time calculator (ETC). It also supports physical unit conversions. It currently works for most ESO instruments for which transmission curves are available through ETC.

Further efforts are required to capture the history of instrumental light paths, e.g., keeping track of filter changes. In the more global context of the Virtual Observatory community efforts are underway to define photometric systems and filter curves in a generally agreed format.

5. WORK IN PROGRESS

The upcoming upgrades to the archive interface will entail many aspects, some of which are listed below:

- producing and publishing science ready products (ADP group): HARPS, GOODS, QC pipeline products, ...
- submission interface for large programs and surveys data products
- meta data mediation utility (`MEx`)
- new archive GUI and content management system
- visual browser
- more previews, instrumental footprints
- efficient regional search (HEALPIX)
- packaging: science - calibration - ancillary data products
- associating calibration frames (calSelector)
- access to filters, transmission curves and photometric standards
- links to publications, proposal database, acknowledgements
- integral field unit/3D support: serve cutouts
- publishing registry to International VO Alliance
- integration of desktop applications via VO messaging interface (PLASTIC)

An initial release is scheduled in Q3 2007 will be followed by gradual updates throughout 2008. By then the full new feature set will become available.

ACKNOWLEDGMENTS

This work is supported by the EURO-VO Technology Centre, a project awarded financial support by the European Commission through contract VOTech 011892 under the 6th Framework Program for research, technological development and demonstration activities. This article summarises efforts by the Virtual Observatory Systems Dept. of the Data Management and Operations Division at ESO, in particular F. Chéreau, N. Delmotte, P. Harrison, P. Padovani, J.C. Malapert, J. Retzlaff, B. Rino, C. Rite, P. Rosati, R. Slikjhuis and A. Wicenec. Thanks to P. Ballester for his support in accessing instrumental transmission curves. Further thanks to the Centre de Données astronomiques de Strasbourg for providing their physical unit conversion library.

GLOSSARY

Abbreviation	Meaning
ADP	Advanced Data Product group
ETC	Exposure Time Calculator
FITS	Flexible Image Transport System
GOODS	Great Observatories Origins Deep Survey
GUI	Graphical User Interface
HARPS	High Accuracy Radial velocity Planet Searcher
HEALPIX	Hierar. Equal Area isoLatitude Pixelization
IVOA	International Virtual Observatory Alliance
MEx	Metadata Extraction utility
PLASTIC	protocol for desktop application interoperability
QC	Quality Control(led data)
SED	Spectral Energy Distribution
SIA	Simple Image Access protocol
SSA	Simple Spectral Access protocol
VISTA	Visible and Infrared Survey Telescope for Astron.
VLT	Very Large Telescope
VST	VLT Survey Telescope
WCS	World Coordinate System
zCOSMOS	spectroscopic redshift survey

[1] `http://archive.eso.org/apps/mag2flux/`

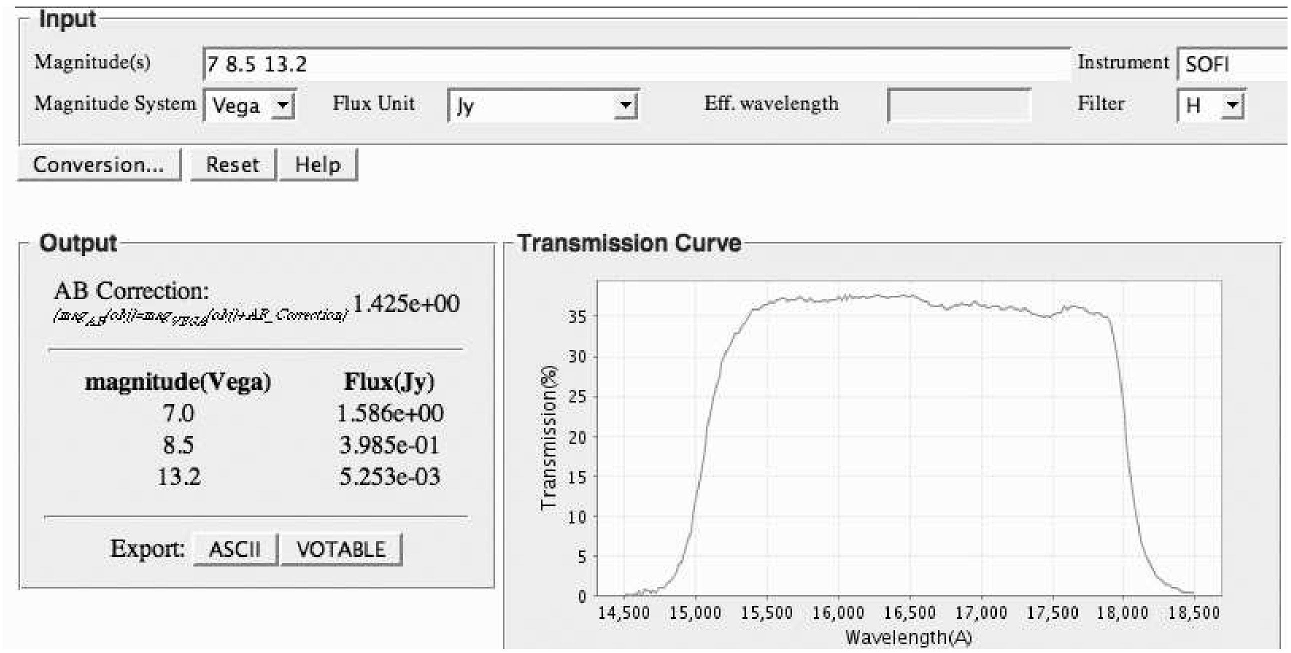

Figure 5. Magnitude to flux conversion `Mag2Flux` *web form.*

REFERENCES

Bonnarel F., Boch T., et al. 2007, Footprint Overlay Specification using current VO Standards: VOTABLE, SIA and STC, http://www.ivoa.net/Documents/

Delmotte N., Dolensky M., et al. 2006, ESO Messenger, 125, 41

Delmotte N., Dolensky M., et al. 2006, ASP Conf. Ser., 351, 690

Leoni M., Dolensky M., et al. 2006, ASP Conf. Ser., 351, 414

Ochsenbein F., et al. 2004, VOTable Format Definition, http://www.ivoa.net/Documents/latest/VOT.html

Slijkhuis R., Delmotte N., et al. 2006, ASP Conf. Ser., 351, 425

Tody D., Plante R., et al. 2004, Simple Image Access Protocol, http://www.ivoa.net/Documents/latest/SIA.html

Tody D., Dolensky M., et al. 2007, Simple Spectral Access Protocol, http://www.ivoa.net/Documents/latest/SSA.html