ESO Science Archive: 1D spectra publishing process

*ESO archive evolving from raw to science-ready*
Outline

- Introduction:
  - ESO science archive evolution
  - Highlighting problems/solutions

- ESO Science Data Product standard for 1D spectra

- Generating/publishing SDP-compliant spectra?

- SDP VO compliance & usability

- Conclusions
Introduction (1)

- Evolving from raw to science-ready
  - ESO started as a RAW data archive

![Bar chart showing ESO vs HST/ECF retrievals comparison in 2005](image)

**ESO vs HST/ECF retrievals: comparison in 2005**
Introduction (2)

Existing science/advanced data products

Advanced Data Products (ADP)

- FEROS spectroscopic time-series data releases (August 2010)
- LABOCA ECDFS Submillimetre Survey data release (March 2010)
- VLT/SAAC HUDF Ks-band imaging data release (March 2010)
- GOODS/VIMOS Spectroscopy final data release (December 2009)
- GOODS/VIMOS Imaging data release (April 2009)
- zCOSMOS second data release DR2 (October 2008)
- GaBoDS/WFI data release update (March 2008)
- ESO/MVM data reduction software release (February 2008)
- Monitor NGC 2547/WFI data release (January 2008)
- GOODS/FORS2 final data release (October 2007)
- GOODS/ISAAC final data release (September 2007)
- FEROS/HARPS pipeline processed data query form
- UVES reprocessed data and other Advanced Data Products
- 30 Doradus/WFI data release (December 2006)
- XMM/WFI Survey Release (September 2005)

Other Public Datasets

ESO Data Releases
- VISTA Science Verification
- APEX-2A and LABOCA Science Verification
- WFI@2.2m Science Verification Data
- The NTT SUSI Deep Field
- The SOFI Infrared images of the NTT Deep Field
- QSO survey: large public programme (VLT UVES)
- FORS1 Supernova 1999em data
- FORS1 - ESSENCE high-z Type Ia Supernovae (April 2005)
- NACO Titan April 2004 release
- GOODS has lots of ESO VLT (and other data) available worldwide

Other Data Releases
- HST Data Releases are available here.
## Archive user interfaces

<table>
<thead>
<tr>
<th>Category</th>
<th>Query Forms</th>
<th>Data collection</th>
<th>Data Type</th>
<th>Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPO Raw Data</td>
<td>Raw data query form (all instruments)</td>
<td>All ESO raw data</td>
<td>Various</td>
<td>Many La Silla Paranal Instruments</td>
</tr>
<tr>
<td></td>
<td>Instrument specific query forms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Direct retrieval of raw data by file name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPO Data Products</td>
<td>Phase 3 main query form</td>
<td>Phase 3 Data Products (ESO Public Surveys)</td>
<td>Currently, Imaging</td>
<td>Currently, VISTA/VIRCAM</td>
</tr>
<tr>
<td></td>
<td>Phase 3 Imaging query form</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase 3 instrument specific query form</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalogue Facility query interface</td>
<td>Phase 3 Catalogues [ESO User Portal authentication required also when browsing]</td>
<td>Catalogues</td>
<td>Currently, VISTA/VIRCAM</td>
<td></td>
</tr>
<tr>
<td>Advanced Data Products query form</td>
<td>GOODS (C.Cesarsky) Imaging, Spectroscopy</td>
<td></td>
<td>FORS2/ISAAC/VIMOS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>zCOSMOS (S.Lilly) Spectroscopy</td>
<td></td>
<td>VIMOS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observation of Cort astroseismologically-selected HD stars (E.Poretti)</td>
<td>Spectroscopy</td>
<td>FEROS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>time series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UVES reprocessed</td>
<td>Spectroscopy</td>
<td>UVES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time-domain survey of NGC 2547 (S.Aigrain)</td>
<td>Imaging</td>
<td>FEROS</td>
<td></td>
</tr>
<tr>
<td>FEROS/HARPS pipeline processed data query form</td>
<td>FEROS/HARPS pipeline processed data</td>
<td>Spectroscopy</td>
<td>FEROS, HARPS</td>
<td></td>
</tr>
<tr>
<td>Science Verification, Commissioning, EIS, etc.</td>
<td>Full list of available data packages</td>
<td>Various</td>
<td>Many</td>
<td></td>
</tr>
<tr>
<td>APEX Quick Look Products</td>
<td>APEX query form</td>
<td>APEX</td>
<td>Heterodyne, Bolometer</td>
<td>APEX-2A, LABOCA, SABOCA, SHeFI</td>
</tr>
</tbody>
</table>
Introduction (4)

■ Archive VO Interfaces (by VOS Dept.)
  ➢ Some VO protocols were introduced:
    • SIAP of DSS images
    • SIAP of (few) selected datasets
    • SSAP of UVES and HARPS pipeline-processed data
  ➢ VirGO archive browser (based on Stellarium)

■ Current status of VO Interfaces:
  ➢ VirGO (no longer supported)
  ➢ Existing SIAP and SSAP run unmaintained (no new data)
Problem Domain

■ Archive systems & user interfaces
  ➢ Plethora of non-integrated archive subsystems
  ➢ Different metadata (names, units, formats, semantics)
  ➢ Different serialization (some FITS, some TAR)
  ➢ Different database tables
  ➢ Specialised user interfaces
  ➢ Maintenance burden
  ➢ Difficulty to grow
  ➢ Impossibility of providing higher-level services

■ On going effort to improve the situation
  ➢ Solutions identified (Archive Roadmap, 2012)
  ➢ Resource-limited effort is on going
Solution domain: stepping stones (1/2)

- Reunification of all archived data products via:
  - Common **Archive Storage** system
    - Has always been common, currently hard drive-based (NGAS, 2001)
  - Unified **Access Control** system (ACE, 2010)
    - Migration of all existing systems to ACE completed in 2012
  - **Phase 3 Science Data Products** standard for Imaging (SDP, 2010)
    - ref. J.Retzlaff’s presentation
  - **CalSelector** (2011)
    - Automatic association of calibration files to science data
  - **Ingestion of External Data Products [EDP]** (Phase3, 2011+)
    - Currently: public surveys. Later: large programs, and potentially PIs of other programs willing to provide their data
  - Phase 3 SDP extended to **1D spectra** (2012)
Solution domain: stepping stones (2/2)

- Phase 3 SDP was promoted to **ESO standard** (2012)
  - Most notably, Instrument and Software divisions (pipelines!)

- Management decision to **generate in-house & publish pipeline-process data** of selected instruments (2012/2013)
  - Internal Data Products [IDP] will complement the EDPs
  - Starting with UVES-ECHELLE (2013, on-going)
  - **Adopting SDP format!**
    - Unified ingestion, Common metadata

- Porting all archive subsytems from **SOLARIS to Linux**
  - Deployment of Request Handler next week will complete the porting of all the main archive components.
  - Completing transition within 2013.
Current Phase 3 Archive Status

- As of now, Phase3 archive contains Imaging only
- Starting 2013 two integrated input channels:
  - EDP: External Data Products
  - IDP: Internal Data Products
- During 2013, ingestion of 1D spectra:
  - Spectral EDPs from Spectroscopic Public Surveys
    - PESSTO, S.Smartt (EFOSC2/Optical, SOFI/NIR)
    - GAIAESO, G.Gilmore+S.Randich (FLAMES)
  - Spectral IDPs
    - Starting with UVES-ECHELLE (all data since 2000)
    - Next: GIRAFFE, XSHOOTER
Part 2

Science Data Product format for 1D spectra
Science Data Product format for 1D spectra

- SDP extended to 1D spectra in 2012
  - After an important study-phase
    - Inputs from various groups (ASG, QC, Science-grade DP, etc)
    - ESO and various ground and space-based instrument data formats were compared.
    - Requirement: spectrum, errors, sky bg in the same file
    - Future proof: support for non equally-sampled arrays
    - Willingness to follow VO standard played a role.

- The binary table format was chosen.
  - Based on the IVOA SpectrumDM 1.0 FITS format
SDP format for 1D spectra

- Issues
  - IRAF spectroscopic tasks (e.g., onedspec/splot) not capable of handling VO/SDP format
  - ESO instrument calibration pipelines do not (yet) support SDP format
  - Psychologically difficult to move away from the “ground-based 1d IMAGE spectral data format”
SDP format for 1D spectra

Good news:

- IRAF support expected!
  - By next IVOA (Sep 2013), Mike Fitzpatrick has promised to deliver an IRAF external package able to read the VO format
  - Later to be part of IRAF kernel

- SDP is now an ESO-wide standard
  - Instrument pipeline will implement the SDP format
  - What to do in the mean time? (see next slide)
Generating 1D spectra in SDP format

Currently ESO pipelines do not support SDP

- PIs of Public Surveys need to implement the SDP format by themselves (validation tool comes handy)

- Internally generated pipeline-processed data need to be converted to SDP (conversion tool)
  - Conversion tool being developed for UVES-ECHELLE
  - User Requirement Document provides recipes for the mapping of the UVES-ECHELLE pipeline products to the SDP format (both keywords and data)
  - Tool developed in Python (PyFITS)
  - Similar effort started for FLAMES/GIRAFFE and HARPS data
  - Additional GIRAFFE complexity: N spectra packed in 1 image
Publishing 1D spectra

Effort on going to:

- Extend the metadata extraction process to 1D spectra
- Extend Phase 3 query forms to 1D spectra

Publishing systems ready by Q3 2013
Science Data Product format for 1D spectra

What is the level of **VO compliance & usability** of a SDP-compliant spectrum?

Luigi’s question: What is needed to make a VO spectrum compliant with the ESO SDP format?
VO compliance & usability

Mandatory keywords

Almost all VO mandatory keywords are present and mandatory as well in the SDP standard;

- VOCLASS, VOPUB, TITLE, OBJECT, RA, DEC, TMID, TELAPSE, SPEC_VAL, SPEC_BW, {TTYPE, TUTYP, TUNIT, TUCD} for FLUX and WAVE

Only exception: TDMINn/TDMAXn absent for n>1 (“TDMIN/TDMAX applies to TTYPE1 only“)

- How bad is this? VO tools are quite permissive and would probably be able to still work with SDP spectra.
- Luigi’s happy: providing those extra values is not a problem.
VO compliance & usability

Recommended VO keywords

- Missing recommended VO keywords:
  - DS_IDPUB
  - RA/DEC_TARG (confusion: target vs telescope pointing)
  - STAT_ERR (same units as flux)
  - TIME UCD and Unit
  - TIME/SPEC/FLUX-SDIM (dimensional analysis, vospec?)

- Different names:
  - SDP uses REFERENC (instead of VOREF): bibcode or doi
  - MID-OBS (TSTART)
  - MID-END (TSTOP)
  - FLUXERR [%] (SYS_ERR, which units? not specified in VO)

- Present as optional keywords in SDP:
  - SPEC_ERR, SPEC_SYE
### VO compliance & usability

<table>
<thead>
<tr>
<th>VO Optional kw name</th>
<th>Default VO value if absent</th>
<th>SDP standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLUX_CAL</td>
<td>CALIBRATED</td>
<td>Absent (Misnamed/FLUXCAL)</td>
</tr>
<tr>
<td>SPEC_CAL</td>
<td>CALIBRATED</td>
<td>Absent</td>
</tr>
<tr>
<td>SKY_CAL</td>
<td>CALIBRATED</td>
<td>Absent</td>
</tr>
<tr>
<td>TIME_CAL</td>
<td>CALIBRATED</td>
<td>Absent</td>
</tr>
<tr>
<td>TSTART</td>
<td>UNKNOWN</td>
<td>Absent</td>
</tr>
<tr>
<td>TSTOP</td>
<td>UNKNOWN</td>
<td>Absent</td>
</tr>
<tr>
<td>VOREF</td>
<td>UNKNOWN</td>
<td>Absent (Misnamed REFERENC)</td>
</tr>
<tr>
<td>SPECSYS</td>
<td>TOPOCENTER</td>
<td>Mandatory</td>
</tr>
</tbody>
</table>

- VO tools dealing with SDP data will assume potentially wrong values.
  
  - Example: Phase3 offers a reduced but **not flux-calibrated** spectrum: VO tool does **not** read FLUXCAL, and assumes **CALIBRATED**.
**VO compliance & usability**

<table>
<thead>
<tr>
<th>VO TTYPEs</th>
<th>SDP TTYPEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAVE/FREQ/ENER</td>
<td>WAVE</td>
</tr>
<tr>
<td>FLUX</td>
<td>FLUX</td>
</tr>
<tr>
<td>ERR</td>
<td>ERR</td>
</tr>
<tr>
<td>BGFLUX</td>
<td>SKYBACK</td>
</tr>
<tr>
<td>QUALITY</td>
<td>QUAL</td>
</tr>
<tr>
<td>BG_ERR</td>
<td>-</td>
</tr>
<tr>
<td>BGQUAL</td>
<td>-</td>
</tr>
</tbody>
</table>

- VO tools will not find SKYBACK and QUAL fields, but could recognise them by the TUTYPn.
- This is the main drawback for (Luigi’s) data written in VO-compliant form.
Conclusions

ESO has put in place a process to:

- Get the archive content to science-ready status as much as possible: EDP+IDP via Phase 3
  - This coming Saturday, **June 15**, Phase 3 will start receiving PESSTO and GAIAESO spectra! Will be published in Q3.
- Unify infrastructure to offer better services (SDP is the first fundamental step in the right direction)
- Generate preview images/spectra for all the scientific products of Phase 3 (on-going, J.Haase ESA/ESO)
- Have instrument calibration pipeline supporting SDP

Only then it will be possible to start building a more advanced infrastructure and services

- improved data model (a la CAOM [CADC]), improved scientific services (VO protocols, footprints, cutouts, etc)