Instrumental Provenance of ESO Archival Data

A. Delgado, N. Delmotte, M. Vuong

Archive Department

European Southern Observatory, Garching bei München, Germany

Goal and Scope

The ESO Archive [1] contains one of the largest collections of ground-based astronomical data in the world that come from a wide variety of telescopes in Chile and other sources and is evolving to become a full research facility capable of optimizing the scientific return from the data.

In the Virtual Observatory (VO) era it is critical that the holdings of the ESO Archive are made available to the user community via the usage of VO standards. This requires among other things a good knowledge of the instrumental provenance of the data and the identification of metadata from the otherwise inhomogeneous collection to form a VO layer and a provenance database.

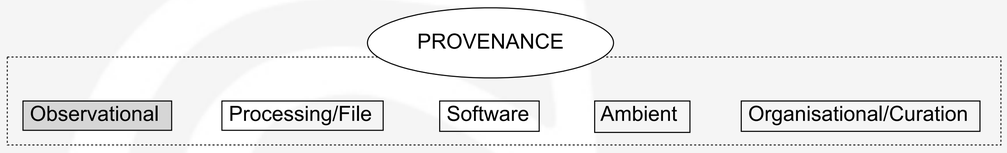
The specifications and transmission curves for all optical elements present in the archived observations are being collected for making them available to the astronomical community. As an example, this homogenisation work makes possible the mining of the ESO Archive looking for objects observed using specific filters that allow the creation of new outreach images.

The goal of this project is to allow scientific access to ESO Archive and hence the query forms functionalities should be enhanced and scientifically oriented, always keeping in mind the astronomers' needs and the VO standards, data models and applications [3]. Also the rapid evolution of the World Wide Web pushes for new functionalities and services.

This project is based only on optical instruments at the present time.

There are various aspects of the astronomical provenance that are further described on Santander et al. [8].

This poster focuses on **Observational Provenance** which describes the configuration and properties of the system(s) used to obtain the data, that is (array of) telescopes, (multi-)chip cameras, filters, grisms... It also includes total transmission curves, instrument description and modes, sensor description, software interfaces etc.



Some example use cases for the instrumental provenance DB would be:

- Spectral Energy Distribution (SED) builder which relies on the availability of the total transmission curves associated with the observations.
- Proposal preparation for astronomers no familiar with ESO instrumentation, to help identifying the instrumental setup that best matches their science cases by querying on unified observational provenance metadata such as spectral coverage, field of view, pixel scale etc.
- Technical query forms of optical elements.

One of the most important informations that can be provided to an astronomer is the total transmission curves related to the scientific observations. For this reason, part of the project is focused on gathering the specifications and transmissions curves of all systems which will set up the basis of the instrumental provenance database (DB).

Methodology

Characterization of Observations and Light Path

A good knowledge of the instruments is very useful for understanding the metadata archived in the ESO DBs. For this reason, all instruments' manuals of different observing periods were collected and consulted for comprehending as much as possible the instruments operational modes [4] [5].
The next step was to identify inconsistencies checking the metadata ingested in FITS keywords repository [2] and the table which contains a selection of the most relevant keywords for the Archive query forms.
Also as a result of this analysis, diagrams 1 and 2 were outlined.

Filter Mappings and Values

One of the results of this analysis showed that some of the keywords values stored in DBs are very inhomogeneous. For instance SUSI, a decommissioned instrument placed at La Silla, has more than 322 distinct strings containing the filters used during the observations, but there were only ~77 filters available for its setup (~53 for SUSI and ~24 for SUSI2). Another example is WFI which has ~46 filters but the number of distinct strings for them in the DB is ~129. The first goal achieved was to collect all the possible strings that were referring to the same metadata values for creating a table containing the mappings. VLT filter name values are more controlled but still some mappings are required.

The absence of unique names for designating each filter was constraining the ability of the Archive query forms for searching by filter name/wavelength. At the present time, most of the frames are reachable querying by filter, but there is a percentage of frames that are left out due to filter name inhomogeneity. This non-unique naming were principally found in La Silla filter's data so in the holdings of earlier ESO instrumentation. The lack of filter information for some frames was also detected and recovered using the keywords repository [2].

In addition, during this detailed examination of the values we could check the consistency of other keyword values such as the observation technique values (i.e. image, spectrum, polarimetry etc.) and the mode's values. So the aim is to find whether these values are referring to one of the observation types described in the manuals but using a distinct string or either these frames are tests or wrongly tagged etc.

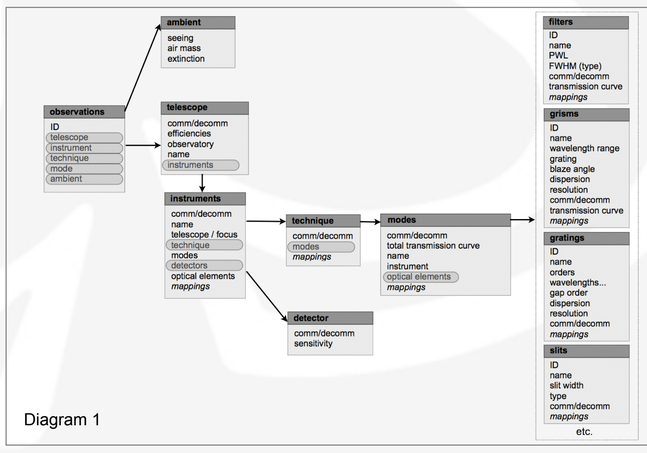


Diagram 1

➡ Characterization of ESO Observations *

The diagram shows how an observation can be characterized from the point of view of the instrumental provenance information.

This simplified model shows the relationships among the different elements at play when performing an observation.

Also the list of specifications and information to take into consideration are given for each element.

The nexus among the elements and the observation are highlighted.

* Interferometry and APEX are not considered in the diagram

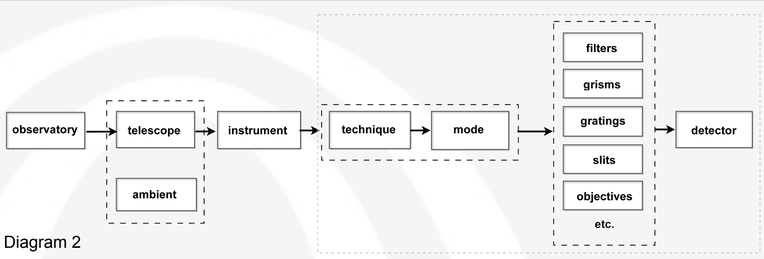


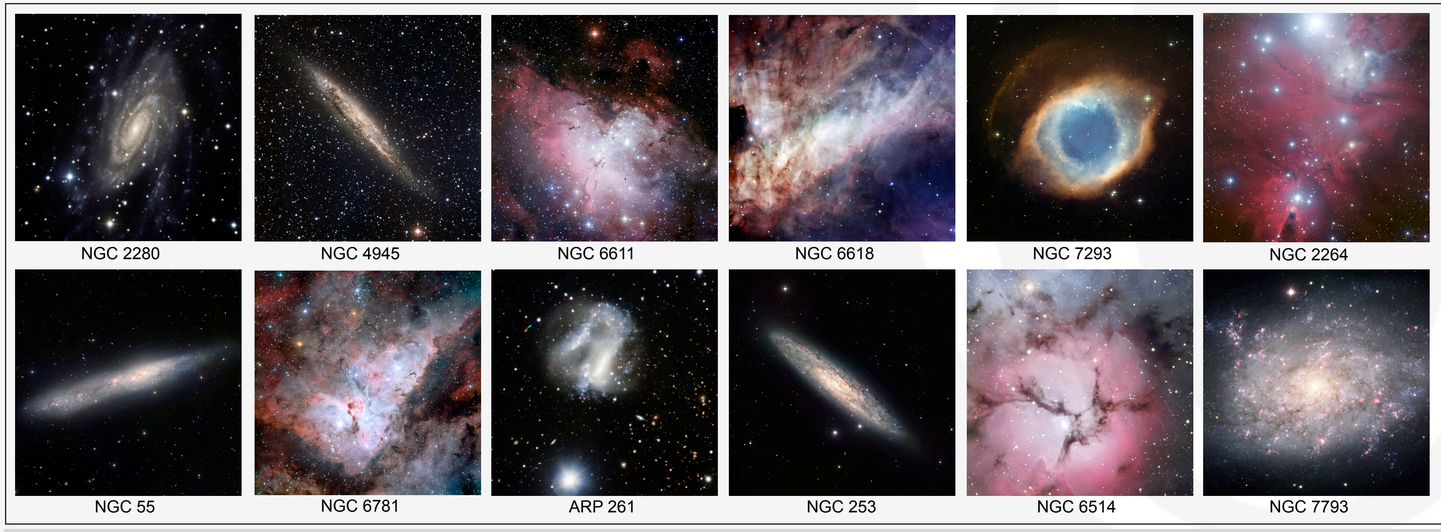
Diagram 2

➡ Data Flow / Light Path *

The total transmission curve associated to an observation is computed as the convolution of the transmission curves of each element crossed by the light coming from outside of the atmosphere.

* Interferometry and APEX are not considered in the diagram

Filters Mappings Applications



Thanks to the filter string homogenization work explained above, it is now possible to get a better characterisation of the Archive holdings. A collaboration has been established with the ESO educational and Public Outreach Department (ePOD) [6] to help mining the Archive for objects which have the needed frames for creating new press release (PR) images.

ePOD provided various lists of objects and set several constraints to the survey:

- The number of science frames per object should be more than three.
- The frames should have been acquired in specific filters, e.g. broad bands (B, V, I..) and Hα, OIII, NII, HeI, HeII etc.
- A simple weighting model was introduced to consider the different importance of the filters for the creation of PR images.

Based on ePOD input, a ranking of the objects more likely to meet all the requirements for the creation of new PR images was produced, including statistics of accumulated frames and exposure times per filter. For the objects selected by ePOD, the list of the frames is provided on request.

Nowadays the search performed in the Archive is part of the foundation of the new image production at ESO. Several images have already been published, based directly or partially on this work. The Archive survey for outreach images is still underway.

Future Work & Conclusions

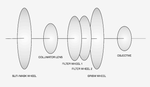
It is time for the design of a data model for keeping track of all instrumental modifications and upgrades undertaken along the years. It will include all the knowledge acquired on the optical elements on and off duty.

In these tables is shown the information that has been gathered from different DBs, manuals, web pages...

Telescopes

- ✓ Site
- ✓ Dates:
  - First light
  - Commissioning / Decommissioning
- ✓ Optical elements
  - Specifications
  - Transmissions curves
- ✓ Instrumentation installed

Instruments



- ✓ Technologies
- ✓ Configurations
- ✓ Dates:
  - Commissioning/Decommissioning
- ✓ Optical Elements
  - Filters
  - Grisms
  - Gratings ....

Optical Elements

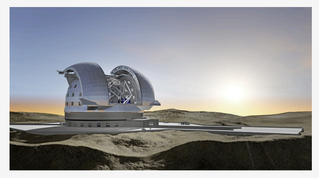
- ✓ Installation date / period
- ✓ Optical Elements:
  - Specifications
  - Transmissions Curves
- ✓ ESO ID
- ✓ Mappings
- ✓ Names
- ✓ Re-calibrations

There are ongoing efforts in the IVOA for releasing an Observational/Provenance data model as well as a Photometry data model for characterising the filters specifications among other data, and the representation of instrumental provenance of ESO archival data will follow recommendation from IVOA

The Instrumental Provenance DBs will give the possibility of creating advanced query forms for allowing the astronomers to get data using scientific parameters in their searches and these features will give the ESO Archive another dimension and more visibility.

The historical information of the observatories will be preserved but what is more important, is the chance of using the archived observations knowing exactly the conditions and the instrumental configuration in the night that they were taken will allow astronomers to make use of them, being confident that they are aware of all the details for use them in their papers.

New instruments at VLT and future facilities like E-ELT should consider including instrumental provenance of observations as an essential element of their archive facilities to keep track of the modifications and upgrades since the cost of the recovery of provenance information after the fact is quite high.



References:

[1] http://archive.eso.org
[2] Vuong et al. 2008 "Applications of the ESO metadata database" Proc. SPIE Vol 7016, 70161M.
[3] http://www.ivoa.net
[4] La Silla Instrumentation: http://www.eso.org/sci/facilities/lasilla/instruments
[5] Paranal Instrumentation: http://www.eso.org/sci/facilities/paranal/instruments
[6] http://www.eso.org/public/outreach
[7] http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaDataModel
[8] Santander et al., "Data Provenace: Use Cases for the ESO archive, and interactions with the VO", Proceedings of the ADASS XIX, ASP Series (to be published)

Acknowledgements:

This work is part of the Spanish in-kind contribution to ESO.

This development has been supported from the Spanish MEC through grant CAC-2006-47.