

ROUND TABLE DISCUSSION OF THE SESSION: “SPECTROSCOPIC SURVEYS”

Moderators: L.Colina & P.Rosati

Introduction: the current context for spectroscopic surveys entails four different types of projects:

- *Spectroscopic survey facilities* (e.g.: SDSS, GALEX). They are characterized by:
 - large volume (~million spectra) of well characterized spectra
 - good user-oriented tools
 - being defined as community service from day one (this is not the rule yet)
- *Large/Key science goal-oriented projects* (e.g.: Spitzer, VLT, Herschel). They will produce specific outputs, not clearly oriented for the most general user. At the same time only a small fraction of total data will be available for the community at large
- *Multi-wavelength spectroscopic surveys* (e.g.: COSMOS). They have a clear astrophysical goal: the generation of broad-band Spectral Energy Distributions (SEDs) for large sample of sources. The main issues implied by this approach are: the cross-calibration among different instruments, and the uniformity of the data (spectral resolutions, apertures, etc.)
- *future spectroscopic facilities/surveys* (e.g.: VLT, ALMA, EELT), which will impose new challenges both in terms of data holding size and on a stronger emphasis toward 3-D spectroscopy

In the light of these developments, the questions that we would like to be addressed during the discussions are:

- What are the challenges of combining different multi-wavelength surveys and heterogeneous spectroscopic data sets in the VO? (cross-calibration, aperture effects, spectral resolution, etc.)
- While services to deal with homogeneous surveys already exist, are there missing tools/functionalitys?
- Are the requirements on data characterization (detailed pedigree) and “curation” currently rigorous enough to combine spectrophotometric information from different data sets?

- How far are we from having a seamless treatment of spectra and SEDs from current surveys (in combination with standard libraries of models and templates)
- Does it make sense to have data/metadata requirements working for all possible scientific cases or should there be requirements specific to science goals?
- Should VO build services for browsing/discover the large variety of 3D spec.data sets which will dominate future archives

McDowell: The characterization model provides the theoretical infrastructure to combine inhomogeneous data from multiwavelength surveys, but the software is not ready yet.

Cerviño: The VO is nowadays good enough for working with one galaxy or one object at the time. However, it is still difficult to achieve a full control on data characterization if one wishes to work with a list of objects.

Genova: However, the parameters in the characterization models reflect requirements gathered from the community. It’s not a closed list, anyhow.

Chilligarian: If we forget for a moment aperture effects, there are already VO tools which properly combine data accessed through SSA servers.

Cropper: The underlying basic issue is quality. Data quality assurance and characterization it is more the job of individual projects. It should not be a job of the VO to check that the quality flags have been properly set. However, it is a job of the VO to ensure that data quality information is propagated correctly, once it is assumed that the job has been properly done *a priori*.

McDowell: We were told by the IVOA not to be the data police; we can be the data reviewers, though.

Piskunov: There are three steps in achieving data quality:

- data providers have to remove as many artifacts as possible
- data providers should assess the quality of their data and propagate the information

- VO should present the data quality in a reasonable and understandable way

Guainazzi: The main task of the VO in the field of data quality is two-fold

- provide an agreed framework to express data quality in an homogeneous way
- ensure that proper dealing of statistical and systematic uncertainties in VO applications receive the highest priority

Prugniel: The characterization model exists, but we need to see it implemented. This is *the* priority now.

Kerber: We need in the VO to be more rigorous in the propagation of errors. All information needs to be transferred in the meta-data. The VO needs to advertise good practices in data quality.

Genova: If we believe that the characterization model is still incomplete, we need the list of parameters that would be required. We may gather a Tiger Team, to ensure that we have a proper semantics.

Freudling: Waiting for the best to come, the priority for the data providers is to start implementing the requirements as soon as possible. One is dealing here with an intrinsically iterative process.

Richards: More than from the point of view of data provider, we should start from science user cases. Implementing is more important now than fiddling on the models.

Bonnarel: There will be couple of talks in the afternoon where some examples of implementation of the characterization model will be shown. This initial steps show that the model can work, and that we can learn something from these implementations.

Genova: The EURO-VO DCA data publishing Workshop at ESAC in June 2007 will be another opportunity to gather feedback from the community.

Tody: Publishing data in the VO is a mediation process. You do not need to understand the VO before, you do not need to adapt the intrinsic structure of your database to any pre-canned VO structure. It's only during the mediation process that VO models need to be understood. At that point one knows if the model is complete enough to describe your data.

Freudling: However, we are left on our own in many decisions on publishing data in the VO, and we should not.

Foellmi: How long does it take to learn publishing data in the VO?

Genova: The French experience tells that it takes around 2 days of intensive learning.

Rosati: Who is going to make sure that the proper meta-data are provided by the data providers?

Tody: When a service is registered, we plan to have a verification software to determine the level of compliance of a server when it is registered. It's important also to have a feed-back mechanism from the data provider.

Rosati: How far are we from having a seamless treatment of spectra and SED (in combination with standard library and templates)?

McDowell: For the time being we are pushing a simple 1-D spectra model only. The same model should, however, be able to cover SEDs as well. Still, a proper photometry model is missing (this implies that a proper color-correction is still not possible). Work is ongoing as well on spectral associations.

Budavari: A preliminary implementation of filter services has been developed (more later).

Rosati: Is there anything in the model which allows users to make proper PSF matching?

McDowell: The aperture is one of the metadata in the model.

Rosati: Does it really make sense to have requirements, which work for all possible science cases?

Cerviño: For the publication of theoretical models references are surely mandatory. Provenance is also mandatory.

Chiligarian: There is a danger in trying and set omnicomprehensive requirements. One may preempt the development to some scientific usages of the data.