"MULTI-WAVELENGTH ASTRONOMY AND THE VIRTUAL OBSERVATORY" WORKSHOP: CONCLUDING REMARKS

Paolo Padovani

European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany

ABSTRACT

I present my personal remarks on the Workshop "Multi-wavelength Astronomy and the Virtual Observatory", which try to capture the main themes that were discussed. I also outline a set of short term priorities for the Virtual Observatory based on the input received from the multi-wavelength astronomical community.

Key words: Virtual Observatory.

1. INTRODUCTION

The Virtual Observatory (VO) is at a turning point as the VO projects, which are members of the International Virtual Observatory Alliance (IVOA)[1], move from development to operations. After its first phase, the Euro-VO[2] project with its three components, the Data Centre Alliance (DCA), the Technology Centre (TC), and the Facility Centre (FC), needs interaction with, and input from, the astronomical community, to make sure that its needs are properly taken into account.

To this aim, and in the framework of the Astronomical Infrastructure for Data Access (AIDA) project, the Euro-VO organized an international workshop at the European Space Astronomy Centre (ESAC), Madrid, on December 1 - 3, 2008.

The goals of the workshop were to pinpoint the challenges multi-wavelength astronomy is facing and to identify how the unique capabilities intrinsic to the VO concept, such as simultaneous access to different archives, metadata describing data content and quality or tools to, e.g., cross-correlate the various datasets, can meet them. In fact, as new and older ground- and space-based facilities continue to provide multi-wavelength data in a variety of formats and large Surveys are expected to increase the data inflow by orders of magnitude, the VO is becoming more and more important for accessing and handling the exponentially increasing data volume.

[1] http://www.ivoa.net/pub/members/
[2] http://www.euro-vo.org/

To achieve these goals, the workshop brought together experts in multi-wavelength galactic and extragalactic astronomy and scientists and engineers actively involved in the VO initiative. In order to ensure an intense debate and a high level of interaction, a substantial fraction of the Workshop was also dedicated to round-table discussions focused on a variety of topics.

This is a personal, biased perspective of what I think were the most important items coming out of the Workshop. All views expressed here are the author's only and do not reflect those of the Euro-VO project, IVOA, or ESO.

2. VO SCIENCE AND DATA ACCESS

It is clear that science with the VO has become a reality (e.g., Caballero, 2009; Chilingarian, 2009; Solano, 2009; Tedds, 2009), as also illustrated by the (conservative) list of VO-based papers, which is maintained at the Euro-VO site[3]. However, there are still some issues limiting the number of astronomical projects, which can take advantage of the VO.

2.1. Science-ready data

One of the major stumbling blocks towards more VO science that became apparent during the Workshop discussions is the still small number of highly reduced products, the so-called "science-ready" data. These are data, which have been processed at the highest level and require very little or no effort to be used by astronomers for their research. Examples include astrometrically and photometrically calibrated images and fully calibrated (wavelength and flux) spectra.

[3] http://www.euro-vo.org/pub/fc/papers.html

Data production, processing, and quality, are the responsibility of data providers, although the VO should do its best to check that the quality of the meta-data (data about data) is acceptable. However, there is interaction between data providers and the VO, which in Europe is provided by the Euro-VO DCA, which coordinates and assists European data centres to take up VO standards, share best practice, and consolidate operational requirements for VO-enabled tools and systems. Therefore, VO projects should convey the message to data providers that more science ready data are needed.

It is fair to say that the number of science-ready data has been increasing lately, but particularly for ground-based observatories their production is quite challenging for various reasons. These include photometric calibration, the non-homogeneous observing conditions and strategies, and the large number of instruments and modes. Most ground-based efforts so far have been dealing with public surveys, e.g., the NRAO VLA Sky Survey (NVSS)[4], the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST)[5], the Sloan Digital Sky Survey (SDSS)[6] (Kent, 2009), the Great Observatories Origins Deep Survey (GOODS)[7], the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS)[8], the NOAO Deep Wide-Field Survey[9], and the UKIRT InfraRed Deep Sky Surveys (UKIDSS)[10]. Exceptions include the reprocessing of the large majority of Ultraviolet and Visual Echelle Spectrograph (UVES) and High Accuracy Radial velocity Planetary Search (HARPS) spectra at ESO[11]. For space observatories, the task of producing science-ready data from generic archival programs is somewhat easier. On-going efforts include the Hubble Legacy Archive[12], the Infrared Space Observatory Highly Processed Data Products[13], the XMM-Newton Archive Data Products[14], and the Chandra Data Archive[15].

The future looks promising. The ASTRONET Infrastructure Roadmap[16] reports that more than half of the facilities surveyed planning to have a public data archive are committed to publishing datasets to the VO and includes a strong recommendation for data centres to provide science-ready data. Furthermore, all the ESO Public Survey Teams[17], which will use the VLT Survey Telescope (VST) and the Visible and Infrared Survey Telescope for Astronomy (VISTA) are required to provide their reduced data products to the ESO archive at the time of publication. This will provide a huge and valuable addition of VO-compliant catalogues and science-ready data to the ESO science archive, available to the world-wide astronomical community.

2.2. Metadata

Lack of appropriate metadata was also mentioned as something hampering more VO-based science at three levels: first, the current VO access is “simple” (Simple Image Access [SIA] and Simple Spectral Access [SSA]) and therefore often very relevant meta-data are not included. For example, limiting magnitude or flux, signal-to-noise ratio, spatial resolution, temporal coverage, and exposure time are currently not part of SIA; second, even when the metadata are there, some VO tools do not allow pre-selection based on them, with the result that it is not easy to find what one is interested in and users are confused by the large amount of output they obtain when performing a query; third, sometimes it also happens that the metadata are not filled in by the data providers. The VO should make sure that all relevant information is provided and make an effort to check that it is correct.

2.3. Raw data

VO protocols so far have been mostly dealing with reduced data. This is because the idea behind VO tools is that they display an image or a spectrum. If the data are in a raw state this is obviously not possible or meaningful. This presents a problem, however, for researchers interested in finding out if a particular target has ever been observed, which currently requires accessing many and diverse archive interfaces. Large (re-)processing/archive projects might also have similar needs. It is important to note that the inclusion of raw data in the VO requires some work on the data providers’ side to make the relevant metadata accessible.

2.4. ... and more

Finally, other causes, which were mentioned as impeding VO science include community unawareness, tools still not adequate or user friendly, and lack of theoretical models in the VO.

3. VO TOOLS

VO tools have made a huge progress in the past few years (e.g., Allen, 2009; Osuna, 2009). One of the main recent developments, thanks mostly to Euro-VO efforts, has been a protocol (PLASTIC, soon to become SAMP), which allows VO tools to interact with a “click of a mouse” permitting, for example, Aladin to “send” a catalogue to TOPCAT or a spectrum to VOSpec.

[4] http://www.cv.nrao.edu/nvss/
[5] http://sundog.stsci.edu/index.html
[6] http://www.sdss.org/dr7/
[7] http://www.stsci.edu/science/goods/
[8] http://www.cfht.hawaii.edu/Science/CFHTLS/cfhtlsdataretrieval.html
[9] http://archive.noao.edu/ndwfs/index.html
[10] http://surveys.roe.ac.uk/wsa/dbaccess.html
[11] http://archive.eso.org/eso/eso_archive_adp.html
[12] http://hla.stsci.edu/
[13] http://iso.esac.esa.int/ida/hpdp.php
[14] http://xmm.esac.esa.int/external/xmm_products/
[15] http://cxc.harvard.edu/cda/
[16] http://www.astronet-eu.org/-Infrastructure-Roadmap-
[17] http://www.eso.org/sci/observing/policies/PublicSurveys/

There are still, however, some problems and missing features, which were brought up at the Workshop, including:

1. too many tools, which in some cases appear to do very similar things; some of them were also judged to be too complicated;
2. the difficulty (or in some cases impossibility) to carry out VO searches for lists of sources, needed, for example, to produce image cutouts or to access thousands of spectra at the same time;
3. the impossibility to make a quick check on all available data, which is related to the raw data access mentioned above (Sect. 2.3);
4. the lack of a VO standard to describe filter/response curves and zero point, which is needed to perform semi-automatic magnitude to flux conversion; the latter is vital to build Spectral Energy Distributions (SEDs);
5. the lack of a so-called Table Access Protocol, which would be the VO standard to deliver catalogues (this is an issue the IVOA is currently working on);
6. a single entry point to the VO. Users appear to get confused by the many Web pages of the various VO projects, and would appreciate a single top-level "VO portal". This should be the first entry returned when one does a Google search on "Virtual Observatory" (the reader should try it now).

One point to keep in mind is that there will never be a "killer" application, that is a VO tool, which does everything an astronomer wants. This is an impossibility, as every science project is different and no single tool will ever be able to satisfy the needs of them all.

4. SURVEYS

Medium/large surveys have less pressing "standard" VO needs, as very often they collect their own data at various wavelengths (see, e.g., GOODS and COSMOS) and therefore have direct access to them. Moreover, no VO tool will ever be able to compete with specifically designed survey tools. Nevertheless, some surveys might still benefit from the VO for the process of gathering information on pre-existing data for their sources. Moreover, some VO tools to facilitate, e.g., an easy match between imaging and spectral data or automatic generation of catalogues (McCracken, 2009) could also be useful. Large surveys are important in a VO context as they are on the "provider" side. In this respect, they need tools to publish data in the VO, which are currently available. Indeed, the Euro-VO Data Centre Alliance project has already organized two Workshops for data providers[18].

[18] See http://esavo.esac.esa.int/EuroVOWorkshopJune2007/ and http://www.euro-vo.org/dcaworkshop2008/.

During the survey round table discussions it was suggested that it would be good to have a VO service, which could provide an answer to the following question: was a position in the sky ever observed in a given band and, if so, is there a detection at a given statistical significance? This can be broken down into three possibilities: a) a catalogue entry is available, in which case the answer is relatively easy; b) reduced data are available, in which case the answer is difficult to give but still possible, as it would require some sort of on-the-fly extraction; c) only raw data are present, in which case it would be extremely hard to get an answer, as it would require on-the-fly data reduction, perhaps easier for space-based observatories but much more complex for ground-based ones.

5. VO TOP PRIORITIES

Based on the input obtained at the Workshops and the round table discussions, I think the near future top priorities for the VO as far as the multi-wavelength astronomical community is concerned are pretty well defined. These include:

1. allowing for searches for list of sources (or making them easier);
2. defining a VO standard for catalogues;
3. exposing to VO searches all available metadata; if more are needed, these should be added to VO standards and made compulsory for data providers;
4. providing some simple ways to build SEDs on-the-fly, as a start at least when catalogue data and science-ready spectra are available;
5. providing a top level "VO portal";
6. providing more information/documentation on VO tools;
7. including raw data in VO searches, with the help of data providers.

6. (MY) VISION FOR THE PETABYTE FUTURE

To conclude, I present my vision for a future with ever larger amounts of data.

Larger telescopes and more complex instruments (e.g., Extremely Large Telescopes, Large Synoptic Survey Telescope) will make data reduction (almost) become a thing of the past for the typical astronomer (as will basically be the case with VISTA data, for example). Most astronomers, therefore, will only deal with catalogues and science-ready data. All of these data will need to be well-described and interoperable, otherwise it will be impossible to make sense of them. Data providers not conforming to this model will be quickly weeded out and go extinct, as their data will not be used and funding agencies will stop supporting them. “Natural selection” will also be at work in the case of the many VO tools, which are currently available: usage will decide, which will survive. And the same will apply to astronomers: “dinosaurs” refusing to embrace the new way of doing things will publish less, be less quoted, and will soon disappear from the scene. And perhaps nobody will remember what the VO was, although everybody will be using it!

ACKNOWLEDGMENTS

The Workshop was supported by the European project Euro-VO Astronomical Infrastructure for Data Access (AIDA) funded by the European Commission through contract number 212104 under the 7th Framework Programme. It is a pleasure to thank all Workshop participants for their input and contributions and Mark Allen and Evanthia Hatziminaoglou for useful comments on the manuscript.

REFERENCES

Allen, M. 2009, these proceedings

Caballero, J.A. 2009, these proceedings

Chilingarian, I. 2009, these proceedings

Kent, S. 2009, these proceedings

McCracken, H.J. 2009, these proceedings

Osuna, P. 2009, these proceedings

Solano, E. 2009, these proceedings

Tedds, J. 2009, these proceedings