FUTURE SPECTROSCOPIC SURVEYS AND THE VIRTUAL OBSERVATORY

Mark Cropper

Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, United Kingdom

ABSTRACT

Future spectroscopic surveys will grow from $\sim 10^5$ objects currently to $\sim 10^8$ objects in 10-15 years. These will address major areas of science, in particular the characteristics of Dark Energy, the formation of galaxies in general, and in particular the structure, formation and evolution of our own Galaxy. VO techniques will be essential in the maximal scientific exploitation of these new resources but current VO tools are relatively limited. Future VO tools will be required for analysing and extracting relevant parameters from large datasets from different origins, for data mining and for sample visualisations. They will be most useful if they provide a robust treatment of observational uncertainties, and a quantification of the statistical biases introduced by the analysis and selection functions.

Key words: Standards; Surveys; Methods: data analysis; Methods: statistical; Techniques: spectroscopic.

1. INTRODUCTION

The Virtual Observatory (VO) is beginning to transform the way astronomers access, process and exploit scientific data. Standards have been agreed in terms of data descriptions, units and processes. The usefulness and usability of the tools is increasing, and it is clear that the ability to incorporate data from different sources in a seamless way is allowing new science to be done.

While the VO has made considerable progress in handling imaging and table data, spectroscopic applications have been relatively neglected. Current VO spectroscopic applications tend to be optimised for the visualisation of small datasets. As spectroscopic data resources grow in size, the VO faces new challenges. This expansion is already taking place with several large spectroscopic surveys now producing more than 10^5 spectra.

This paper examines the future of spectroscopic surveys. It starts with a census of the currently approved and leading proposed surveys. It then identifies features in common, and then examines what VO capability might be useful to maximise the scientific return from both the individual surveys and their combination.

2. FUTURE LARGE SURVEYS

Spectroscopic surveys will grow by orders of magnitude in size over the next 10–15 years. This increase will be driven by satellite all-sky surveys, multi-object spectrographs with > 1000 fibres, large integral field spectrographs and by major new sub-mm and radio facilities (ALMA, SKA).

Future large spectral surveys fall into three broad categories: dark energy surveys (baryon acoustic oscillation or BAO surveys), surveys of galaxy formation, and surveys to determine the structure of our Galaxy. Table 1 and 2 identify the major surveys in these categories, both approved and proposed. The criterion adopted here is that the survey should produce observations of $> 10^5$ sources. This is not a complete list, but the tables provides a fair and representative overview of what spectroscopic resources will be brought to bear, the scale of the survey, its general characteristics (wavelength range, resolving power) and when the survey will take place. In one or two cases (such as RAVE), the survey has commenced already, but the main challenge for making the data globally available, for example the via the VO, have yet to be addressed. In each case hyperlink references have been made available to facilitate further investigation: details may evolve as the later surveys are refined.

2.1. Dark Energy surveys

There are several strategies for characterising the dark energy content of the universe, The spectroscopic BAO surveys (Table 1) measure the redshifts of galaxies to obtain distances, which combined with positions on the sky provides the 3-dimensional distribution of galaxies. These

Table 1. Approved and proposed large-scale spectroscopic Dark Energy surveys.

Name	Facility	Total Size	Spectra per exp	Limiting Mag	Spectral Range	Resolving Power	Duration	Status	refs
WiggleZ	AAT+AAOmega	4×10^5 star forming galaxies	400	$r < 22.5$	$370 - 900$ nm	$1 - 2000$	2006–2009	approved	[1,2]
FastSound	Subaru+FMOS	6×10^5 star forming galaxies^a	400	$J < 22$	$0.9 - 1.2\mu$m	~ 500	~ 2008	proposed	[3]
SDSS post-2008	Apache Point 2.5m	10^6 galaxies and Lyα absorbers	640	$g < 20$	$380 - 920$ nm	~ 2000	2008–2012	proposed	[4]
HEXDEX	HET+VIRUS	10^6 Lyα emitting galaxies	36000	$AB < 23.5$	$340 - 570$ nm	~ 850	~ 2010	approved	[5,6]
LAMOST	LAMOST+LRS	10^7 Luminous Red Galaxies	4000	$V \sim 20.5$	$370 - 900$ nm	~ 2000	~ 2008	approved	[7,8]
Subaru/WFMOS	Subaru+FMOS	3×10^6 galaxies $z = 0.5 - 3.3$	4500	$AB < 22.7$	400 nm^b	13500	~ 2012	proposed^c	[9]
ADEPT	ADEPT satellite	10^8 galaxies $z = 1 - 2$	^{d}	^{d}	$\sim 1.3 - 2\mu$m	^{d}	after 2015	proposed^e	[10,11]
SKA	SKA	$10^8 - 10^9$ galaxies $z < 3^f$	^{g}	^{h}	$0.1 - 25$GHz	^{i}	~ 2019	in study	[12]

Notes:

a – via Hα measurements

b – within wavelength range 400nm – 1μm

c – in negotiation

d – details not available due to competitive bid

e – 1 of 3 contenders selected for the JDEM in the NASA Beyond Einstein programme

f – Using HI

g – large numbers of beams will be available

h – limiting sensitivity $\sim 0.01 \times VLA$

i – 10000 spectral channels at a time within this range

References:

[1] http://astronomy.swin.edu.au/wigglez/WiggleZ/Welcome.html

[2] http://arxiv.org/abs/astro-ph/0701876

[3] http://www-utap.phys.s.u-tokyo.ac.jp/ utap/meetings/workshop/hsc2006/English/totani.ppt

[4] http://astro.snu.ac.kr/ sdss/talks/plenary%202/Kron_post-2008_plan.pdf

[5] http://www.as.utexas.edu/hetdex/Hill_MitchellSymposium_mod.pdf

[6] http://www.as.utexas.edu/hetdex/

[7] http://ej.iop.org/links/r_nx_nc4r/atq-vl3V2xGMKyGYav5vpA/chjaa_6_3_001.pdf

[8] http://www.lamost.org/en/

[9] http://www.noao.edu/meetings/subaru/Session-1/Colless.pdf

[10] http://www7.nationalacademies.org/ssb/mtg_2_ADEPT.pdf

[11] http://www.science.doe.gov/hep/HEPAP/Feb2007/HEPAP_Bennett_Feb07.pdf

[12] http://www.skatelescope.org

Table 2. Approved and proposed large-scale spectroscopic Galactic structure and galaxy formation surveys.

Name	Facility	Total Size	Spectra per exp	Limiting Mag	Spectral Range	Resolving Power	Duration	Status	refs
Galactic Structure:									
RAVE	UKSchmidt+6df	10^6 stars	150	$V < 13.5$	$850 - 875$ nm	~ 10000	2003–2010	under way	[1]
SDSS SEGUE	Apache Point 2.5m	2.5×10^5 thick disk/halo stars	640	$g \simeq 20$	$380 - 920$ nm	~ 2000	2005–2008	under way	[2]
ARGOS	AAT+AAOmega	10^5 mostly bulge stars	400	$I_C \sim 17$	$846 - 886^a$ nm	~ 10000	2007–2009	proposed	[3]
SDSS post-2008	Apache Point 2.5m	2×10^6 stars	640	$g \simeq 20$	$380 - 920$ nm	~ 2000	2008–2014	proposed	[4]
LAMOST	LAMOST+LRS	10^8 stars	4000	$V \sim 20.5$	$370 - 900$ nm	~ 2000	~ 2008	approved	[5,6]
Gaia-RVS	Gaia satellite	10^8 stars	^{b}	$V < 17.5$	$847 - 874$ nm	~ 11000	2012–2017	approved^c	[7]
galaxy formation:									
KMOS	VLT+KMOS	^{d}	200×24	$K < 21$	$0.8 - 2.5\mu$m	~ 3500	2011–	approved	[8]
MUSE	VLT+MUSE	^{d}	90000	$I_C \sim 25$	$465 - 930$ nm	~ 3000	2011–	approved	[9]
Herschel	Herschel +SPIRE+PACS	850 hr (SPIRE) +650 hr (PACS)		230 mJy^e	$60 - 670\mu$m	$\sim 300 - 1200^f$	2008 – 2010	approved^g	[10,11]

Notes:

a – also 525–560 nm
b $\sim$ 40 exposures per star
c – ESA 6th Cornerstone Mission
d – survey details unknown
e – in the continuum (SPIRE)
f – SPIRE
g – ESA 4th Cornerstone Mission

References:

[1] http://www.rave-survey.aip.de/rave/pages/project/ProjectDescription_1.jsp
[2] http://www.sdss.org/segue/aboutsegue.html
[3] http://www.physics.usyd.edu.au/ gfl/ARGOS/
[4] http://astro.snu.ac.kr/ sdss/talks/plenary%202/Kron_post-2008_plan.pdf
[5] http://ej.iop.org/links/r_nx_nc4r/atq-vl3V2xGMKyGYav5vpA/chjaa_6_3_001.pdf
[6] http://www.lamost.org/en/
[7] http://www.rssd.esa.int/GAIA/
[8] http://www.eso.org/instruments/kmos/#Science
[9] http://www.eso.org/instruments/muse/
[10] http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=34682
[11] http://www.astro.cf.ac.uk/groups/instrumentation/projects/spire/Public_Index.htm

can be compared to different cosmological models, with different prescriptions of the dark energy content. In particular, the comparison can be made against the predicted spatial frequency spectrum which traces the acoustic oscillation of the baryon content of the universe. The main requirement is for large surveys at moderate spectral resolution.

2.2. Galaxy formation surveys

Galaxy formation surveys (Table 2) aim to examine the formation and assembly of galaxies at high redshift as they undergo mergers and bursts of star formation. Optical studies are then sensitive to Lyα emission, while infrared surveys work at the peak of the spectral energy distribution of the galaxies. Some spatially resolved kinematic information again at moderate spectral resolution is useful for studying mergers and determining redshifts.

2.3. Surveys of our Galaxy

The surveys on the structure of our Galaxy (Table 2) examine the current state and the evolutionary history of the Galaxy (star formation history, chemical enrichment, merger history). The cosmological implications can be determined by examining the chemical and kinematic record preserved in the stellar populations. These can be quantified on a star-by-star basis only for our Galaxy and the Local Group. While these surveys are generally orientated towards understanding how the Galaxy was formed (illuminating the process for galaxy formation in general), many other important aspects of Galactic studies are addressed by such surveys. They generally have somewhat higher spectral resolution.

As the first spectroscopic survey of $> 10^8$ sources, ESA's Gaia-RVS survey of our Galaxy is a particular driver for how the VO services large spectroscopic surveys. Because each object is observed ~ 40 times, Gaia-RVS will produce a total of $\sim 4 \times 10^9$ spectra. Most of the data will be of very low signal-to-noise ratio. The individual observations will also be combined to produce end-of-mission spectra, and some derived products will be available. These will include the radial (and rotational) velocity, and indications on binarity/multiplicity (and if a binary, the binary period). It may not include line strengths, metallicities and gravities. These data products will be VO-enabled by the teams processing the Gaia data. As far as the VO is concerned, there is a significant need and many opportunities for VO-enabled tools to select, examine, combine, analyse and classify this dataset, and to combine it with Gaia photometry and astrometry.

2.4. Other surveys

Besides the above there are several other important large spectroscopic surveys. Examples include the XMM-Newton/Chandra X-ray surveys constructed mainly from pointed observations, and which continue to increase in sky coverage, the AKARI satellite infrared survey, and radio and sub-mm surveys such as those planned by LOFAR and ALMA.

3. FEATURES IN COMMON

Some comments can be made on the common features of these spectroscopic surveys.

Scale: the scale is (by definition) large, ranging from $10^5 - 10^8$ sources, some observed many times. The largest are SKA, Gaia-RVS, ADEPT, LAMOST and Subaru+WFMOS. The exponential growth in survey size is shown in Figure 1.

Spatial connectedness: some surveys are all-sky, such as Gaia; others provide integral field spectroscopy, and hence are closely connected (VLT-MUSE; VLT-KMOS).

Wavelength range: most of the largest spectral surveys are in the optical/infrared, but significantly sized surveys also at other wavelengths (such as XMM-Newton/Chandra). The largest (in the more distant future) will be in the radio, from SKA.

Signal-to-noise ratios: most surveys will be exposed for reasonable S/N ratios, but the biggest (Gaia-RVS, ADEPT, SKA) will have data with a whole range of S/N ratios, the majority being very low.

Data state: All large surveys will provide reduced data in standardised form, *i.e.* calibrated in standardised units (this will be part of the responsibility of the project itself). Looking forward, data will surely be VO-ready. Different strategies will probably be employed for data storage: mainly databases, flat file structures and VOTables. Data to be distributed will in some cases therefore require VO wrapping on access.

4. WHAT CURRENT VO TOOLS PROVIDE

Current spectral VO capabilities include the following packages:

1. *VOSpec*, developed by the ESA-VO Team [1];
2. *SpecView*, developed by Ivo Busko at STSci [2];
3. *SPLAT-VO*, developed by Peter Draper with UK Starlink support [3];
4. *Spectrum Services for the VO* developed by Tamás Budávari and colleagues at JHU [4].

---

[1] see http://esavo.esa.int/vospec/

[2] see http://www.stsci.edu/resources/software_hardware/specview

[3] see http://star-www.dur.ac.uk/p̃draper/splat/splat-vo/splat-vo.html

[4] see http://www.voservices.net/spectrum/

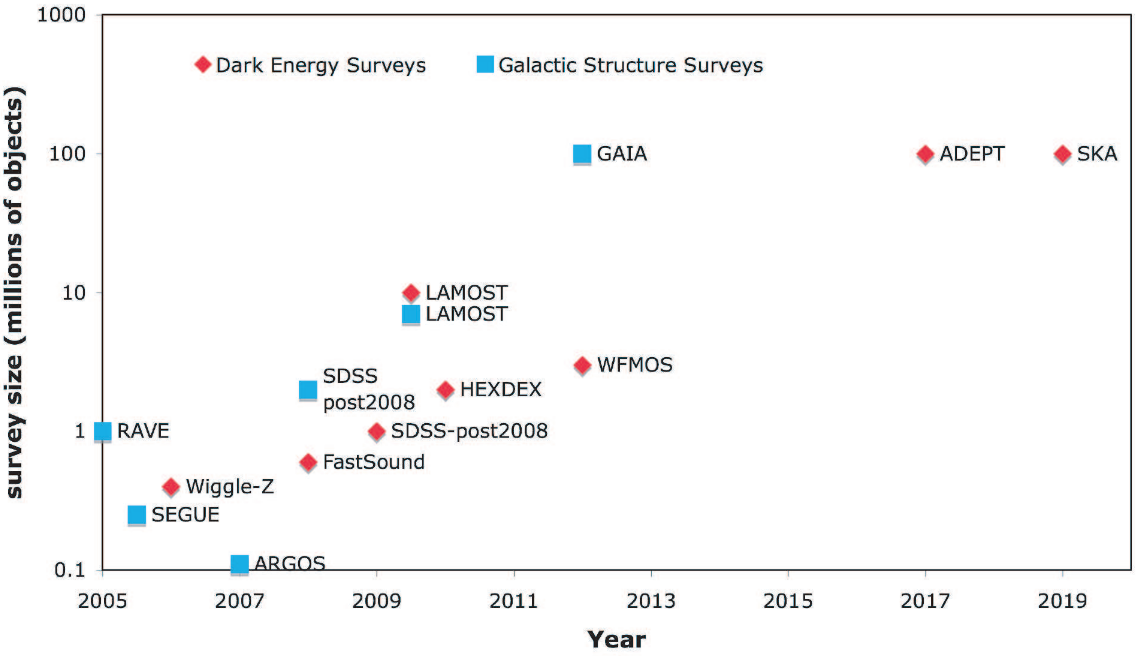

Figure 1. The growth in survey size (from start date of survey) for Dark Energy and Galactic Structure surveys.

These tools appear to be generally similar, but with complementary features and slightly different approaches. They appear to be orientated mostly towards visualisation, together with some data reduction and analysis capabilities. These include dereddening of spectra, flux measurements and fits to generic templates (such as blackbodies) and in some cases to stellar libraries.

In addition, there are other new packages announced at this meeting, such as EZ, GOSSIP (Franzetti et al 2007) and CASSIS (Caux et al 2007).

5. WHAT WE MIGHT LIKE TO DO WITH FUTURE LARGE SURVEYS

Somewhat paradoxically, it is not clear whether VO tools will be used in the reduction and calibration of future large spectroscopic surveys. The scale of these enterprises, combined with the particularities of the data and the availability of a team to ensure that it is processed, makes it more likely that efficient specifically tailored software will be developed. However, certain aspects of the processing within the processing chain will require visualisation or more detailed analysis, so it may be useful to ensure VO compatibility of the data at least at certain stages, so that general VO spectroscopic tools can be applied.

Because the data from future large spectroscopic surveys will already be reduced and calibrated, the main thrust of future VO tools should be in the analysis of large datasets, in the combination of large datasets from different sources, in data mining, and in visualisations of the results of the analyses. These are all areas in which the VO is the obvious way forward.

5.1. Practical Aspects

Practical aspects will affect the success with which VO tools are used for large spectroscopic surveys. Intuitiveness of use will be vital, otherwise they will be used only by specialists. It will also be important to be able to access the content of the surveys with reasonable rapidity: this may drive the methodology of storage, access, processing and product delivery.

The scale of most future large surveys will prohibit the working with and inspection of the majority of individual spectra. Rather, some sort of outlier selection and visualisation will be necessary to deal with particular cases.

It is unlikely that single VO tools can be developed to deal with the range and scale of the possible tasks to be addressed. Some automation of procedures (the building of workflows as in the AstroGrid Workbench[5]) may be the only feasible way forward here. However, this needs to be made less daunting than it currently seems. There are also implications for the interface capabilities of the VO tools that would be plugged into the workflow.

---

[5] see http://www2.astrogrid.org/desktop

5.2. Analysis

A clear requirement for VO tools for large spectroscopic surveys will be to extract parameters from the spectra. These include equivalent widths, line ratios, continuum slopes, velocity shifts or redshifts and rotational velocity or velocity dispersions.

Robust methods will be required for reliable operation on a wide range of spectral data. An example is the difficulty of assigning an adequate continuum fit to M-star spectra. Careful handling of observational errors (uncertainties) is also vital: this is generally a more difficult problem than the observational values themselves. Furthermore, an understanding and recording of the statistical biases (completeness, limits) introduced and propagated by the analysis would be extremely valuable.

It is currently difficult to carry out such analysis tasks in a uniform way across different surveys. The important role of the VO will be to provide this capability across databases in a standardised way, perhaps by adapting methods already developed for individual large surveys. Some tools may already be available for VO adaptation from the processing of the data by the survey consortia.

5.3. Data Mining

Searches for non-standard (‘interesting’) spectra will be important: examples include emission lines, active galactic nuclei, binarity/multiplicity and unusual stellar types.

A major capability that the VO could provide is in determining the connectedness of different datasets. This may be spatial connectedness, such as found in integral field data (MUSE, KMOS) or from mosaiced fields, or it may be in terms of derived parameters such as momentum and energy (phase space connectedness) or such as age, metallicity and redshift. In general this will involve the combination of spectral, photometric and astrometric data from across the electromagnetic spectrum (cross-matching), and may ultimately involve some degree of classification. VO-enabled tools to determine the connectedness, make selections and cuts, visualise the samples and provide statistically rigorous outputs would be extremely useful. Again it may be necessary to achieve this in practise by means of workflows tailored to addressing particular investigations.

6. SUMMARY

Looking forward, there will be spectroscopic surveys of > 10 million spectra within a decade. These surveys will have been justified on scientific grounds within the surveys themselves, but the overall science return will be significantly enhanced by the capability to combine and analyse spectroscopic surveys of all types (and to include astrometry and photometry). The combination of data on such a scale will enable a flowering of new science.

Spectroscopic surveys will require tools to extract information from the spectra and to characterise it in a regularised fashion (using derived parameters). It would be helpful if the VO could examine how this may be done: the generation of workflows in simple intuitive ways (perhaps via a graphical interface) may be the key. VO-provided capability to extract and visualise information from the very large spectral samples, with rigorous statistical treatment of observational uncertainties and statistical biases will be scientifically valuable.

ACKNOWLEDGMENTS

I am grateful to Will O’Mullane, Nic Walton, Pedro Osuna, Peter Draper, Tamas Budavari and Ivo Busko for helpful discussions, and to the SOC for commissioning this review.

REFERENCES

Caux, E., Klotz, A., Vastel, C., Walters, A., 2007, these proceedings.

Franzetti, P. and the Pandora Team, 2007, these proceedings.