SLOAN DIGITAL SKY SURVEY AND MULTIWAVELENGTH ASTRONOMY

Stephen Kent
Fermi National Accelerator Laboratory
for the SDSS collaboration

ABSTRACT

The SDSS and SDSS-II projects (York et al., 2000) have just published their final data archive, consisting of five-band images over 1/4 the sky, a catalog of over 350 million objects, and spectra for 1 million galaxies, 100,000 quasars, and 240,000 stars. The data volume consists of 73 terabytes in files and 18 terabytes in databases.

Of the over 800 papers published so far using SDSS data, over half are analyses by astronomers outside the SDSS collaboration using the public data releases. This talk examines the ways in which the SDSS data are used, including multi-wavelength analyses matching data from the xray to the radio. Several VO interfaces are now available that provide access to the data in parallel with the SDSS-only interfaces. Positive and negative aspects of these interfaces will be discussed. Future surveys that are designed to overlap with the SDSS footprint are also mentioned briefly.

Key words: LaTeX; Virtual Observatory; Sloan; SDSS.

1. SURVEY DESCRIPTION

The SDSS has been truly international in scope - two of the funding agencies and six of the participating institutions are European.

The Sloan Digital Sky Survey (SDSS), recently completed, was a project whose original goal was to map 1/4 of the sky in 5 bands and obtain the redshifts of 1 million galaxies and 100,000 quasars over 5 years of operations during 2000 to 2005 (Fig. 1). The original SDSS did not quite meet all its goals, and as a consequence a second phase of the project, SDSS-II, continued the operations through 2008. In addition to the original goals (now called the Legacy program), two new programs were added to SDSS-II: SEGUE, a project to image 3500 square degrees and measure the spectra and metallicities of 240,000 stars, and Supernovae, a project to obtain well-measured light curves of type I-A supernovae in the redshift range 0.1 to 0.3.

Figure 1. SDSS 2.5-meter telescope.

The imaging stripe layout, plotted in Fig. 2, shows the locations of the legacy area (the filled area in the center), and the SEGUE stripes (a sparse grid of stripes crossing the Galactic plane), plus other stripes obtained for calibrations or to cover other interesting areas of the sky. Each stripe is 2.5 degrees wide.

The legacy spectroscopic survey is essentially complete to predefined limiting magnitudes for both galaxies and quasars (aside from closely-spaced targets that caused fiber collisions). The SEGUE spectroscopic survey sparsely samples the imaging survey in a semi-regular grid of fields (Fig. 3).

2. DATA PRODUCTS AND RELEASES

Before describing the data products, it is interesting to think of how the SDSS data fit into the overall space of all possible types of astronomical data. One of the original motivations for creating the VO was that astronomical data form a complex, multidimensional space, to which access would benefit from having an organiza-

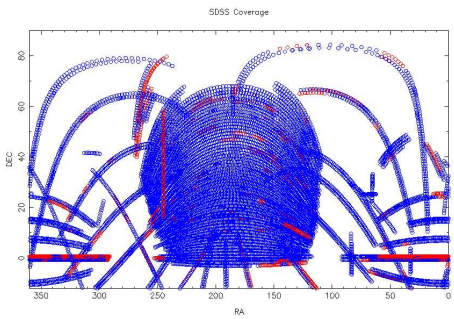


Figure 2. Sky coverage of SDSS imaging survey. The color coding (intended to show data quality) should be ignored.

tional structure like the Virtual Observatory (VO). I like to think of this space as being roughly five dimensional. Two of the dimensions are angular position on the sky, one dimension is depth or limiting magnitude (the two generally corresponding to one another), one dimension is time, and one dimension is frequency, or equivalently energy. In addition to this space, astronomical data can be characterized by five additional parameters, which are the corresonding resolution or accuracy in each dimension that characterize the data. As an added complexity, one could, in principle, characterize according to the type of particle, such as electromagnetic (photons), cosmic ray, neutrino, nuetralino, axion, or graviton, although such complexity is hardly needed at present.

The SDSS only fill a small volume in the above 5 dimension + 5 parameter space. Further, the range of all possible astronomical data, while much larger, also only partially fill this space. In order to join two datasets to conduct a multiwavlength study (the theme of this workshop), then, depending on the type of study, the two datasets must overlap in anywhere between 2 and 4 dimensions (i.e., at least area of sky) and match reasonably close in anywhere from 2 to 5 parameters (e.g., angular resolution) in order to be useful. Characteristic sky coverage and parameters of the SDSS imaging and spectrospic data are presented in Tables 1 and 2.

The SDSS data have been distributed in a series of "data releases" at roughly 1 year intervals. The SDSS-II survey finished in July, 2008, and the final data release (DR7) occurred in October of 2008. Some overall properties of the data release, taken from the SDSS DR7 website, are given in Table 3. The total data volume is approximately 91 terabytes.

The SDSS has developed two principal access mechanisms: the Data Archive Server (DAS) and the Catalog Archive Server (CAS). The DAS provides access to bulk data at flatfiles, generally organized by the way that the data were collected. A form query interface is provided to enable simple queries to locate files of interest. Either rsync or http access can be used to retrieve files. The CAS contains all the object catalogs in the form of three databases with several form and SQL interfaces available.

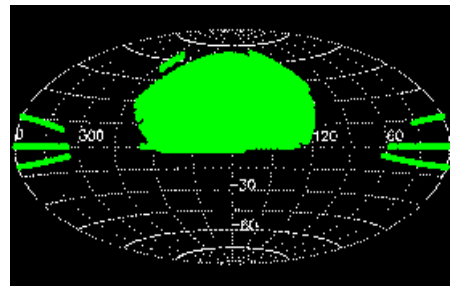


Figure 3. Sky coverage of the Legacy (top) and SEGUE (bottom) spectroscopic surveys. Each dot represents a 2.5 degree diameter field with approximately 600 spectroscopic targets.

Table 1. Imaging Survey Parameters.

Parameter	Value
Resolution $\lambda/\Delta\lambda$	≈ 5
Limiting Mag	$u = 22$ (AB)
	$g = 22$
	$r = 22$
	$i = 21$
	$z = 20$
Saturation	$u = 14$ (AB)
	$g = 14$
	$r = 14$
	$i = 13$
	$z = 12$
Photometric Calibration	1.5%
Median Seeing	1.2 arcsec
Astrometric Calibration	< 100 mas

Table 2. Spectroscopic Survey Parameters.

Parameter	Value
Resolution $\lambda/\Delta\lambda$	≈ 2000
Limiting Magnitude	
Galaxies	$r < 17.8$
QSOs	$i < 19.$
Radial velocity calibration	
Galaxies	30 km/s
Stars	6 km/s

Table 3. Data Release 7 Statistics.

Parameter	Value
Imaging	11,663 sq. deg. unique
	45,000 sq. deg. total
	(including repeats)
Spectroscopy	9380 sq. deg.
Object Catalog	357 million objects
Redshift Catalog	1.6 million objects
Data Volume	
Flat Files	73 TB
Databases	18 TB

The "Casjobs" interface provides a batch-type interface for more demanding and intensive types of queries.

3. DATA USAGE

Through 2008, over 800 papers have been published analyzing SDSS data. Of these, over half have been written by non-SDSS collaborators based on the public data releases.

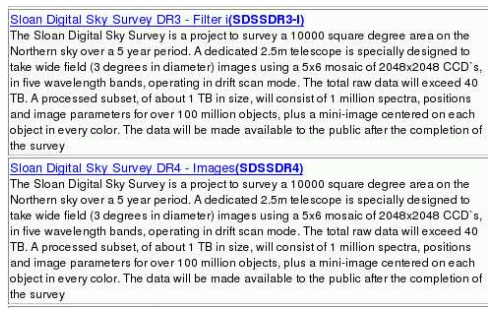
In August 2008, the SDSS held a symposium covering a wide range of science being done using SDSS data. The number of submitted abstracts, 71, is large enough to make some general statements about the types of analyses being done. This analysis gives some insight into what which SDSS "customers" would benefit from the VO.

- About 60% of the abstracts describe work that is "survey science" - analysis of SDSS data only.
- About 30% are "multi-telscope" - combining SDSS data with other data sets.
- About 10% (7 abstracts) describe "multiwavelength" analyses, in which SDSS data are combined with data at other wavelengths, utilizing surveys such as ROSAT, FIRST, GALEX, etc.

4. VO INTERFACES

Since the Virtual Observatory postdates the development of the SDSS archive, VO interfaces have been "grafted on" to the existing archive infrastructure; this work has been conducted as part of the US-based "NVO" project. Thus, while the data are accessible through the VO, the interfaces are somewhat uneven.

As an example, a query on the JHU VO registry for all entries with "SDSS" or "Sloan" returns 195 results. Not all are SDSS data, and not all SDSS data are yet in the registry. Each data release has one or more entries, but they are not consistent from one to the next, and because the later data releases are merely cumulative, a particular piece of data can be covered by multiple entries (Fig. 4).


Sloan Digital Sky Survey DR3 - Filter i(**SDSSDR3-I**)
The Sloan Digital Sky Survey is a project to survey a 10000 square degree area on the Northern sky over a 5 year period. A dedicated 2.5m telescope is specially designed to take wide field (3 degrees in diameter) images using a 5x6 mosaic of 2048x2048 CCD`s, in five wavelength bands, operating in drift scan mode. The total raw data will exceed 40 TB. A processed subset, of about 1 TB in size, will consist of 1 million spectra, positions and image parameters for over 100 million objects, plus a mini-image centered on each object in every color. The data will be made available to the public after the completion of the survey

Sloan Digital Sky Survey DR4 - Images(**SDSSDR4**)
The Sloan Digital Sky Survey is a project to survey a 10000 square degree area on the Northern sky over a 5 year period. A dedicated 2.5m telescope is specially designed to take wide field (3 degrees in diameter) images using a 5x6 mosaic of 2048x2048 CCD`s, in five wavelength bands, operating in drift scan mode. The total raw data will exceed 40 TB. A processed subset, of about 1 TB in size, will consist of 1 million spectra, positions and image parameters for over 100 million objects, plus a mini-image centered on each object in every color. The data will be made available to the public after the completion of the survey

Figure 4. A couple of SDSS entries in the JHU VO registry. Multiple data releases and multiple types of data can lead to confusing registry entries.

Several VO services are now provided, including SIAP, cone search, tabular, Skynode, SSAP, as well as basic http and rsync services mentioned above. These services are access through high level interfaces, including OpenSkyQuery, Datascope, Image Mosaic Service, NVO Spectrum Services, and VO Explorer et al. These interfaces could stand some improvement - e.g., Datascope and VO Spectral Services both take a position and a radius as input, but the format and units are different between the two. Also, Datascope does not return information about spectra of a particular object.

In spite of the above unevenness of access to SDSS data, researchers are able to use them. A search through the NASA Astrophysics Data System for papers that cite both the SDSS and the Virtual Observatory returns at least five papers in recent years, of which three (Shuecker, Böhringer and Voges, 2004; Santos, Mendes de Oliveira, and Sodré, 2007; Gao, Zhang, and Zhao, 2008) qualify as multiwavelength studies. Undoubtedly other similar

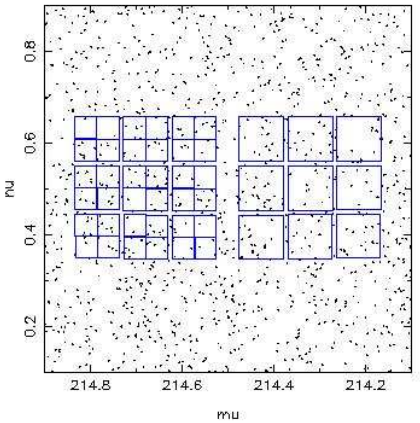


Figure 5. Planning exercise for a future JDEM mission. The figure shows stars with data from SDSS, 2MASS, and USNO-B with an overlay of one possible focal plane layout.

papers exist. I have personally used OpenSkyQuery as a convenient tool to perform cross-matches between SDSS data and other surveys. For example, one exercise was to identify stars with SDSS photometry, 2MASS position, and USNO-B positions as part of planning a future JDEM mission (Fig. 5). Not all exercises were successful - one query failed because of an incorrect reformatting of data for use by the VO.

Integration of the SDSS with the VO has let to several issues, which span both data providers and the VO.

- A long-running survey like SDSS will have multiple reprocessings. Should all versions be kept available to a VO, or just the latest? What happens if someone mirrors early versions of the data?
- A survey like SDSS will also have multiple data releases, some of which also involve reprocessing. Once again, should all version be kept available, or just the latest?
- Data are complex - the SDSS data now comprise three separate surveys, and in the future they will comprise seven!
- An issue more for data providers than for the VO, but relevant nevertheless - SQL database for storing tabular data and metadata are a core component for the VO, but they are not an archival storage format. SDSS is still in search of a long-term steward for its data archive - not everyone can absorb 100 TB with specialized interfaces, and the software components are susceptible to bit rot over time.
- Bulk access to data is still awkward. Command line tools will help alleviate the access problems.
- What role will the VO play in future surveys? In particular, could high-level VO interfaces replace survey-specific interfaces? Probably not, but they might replace some.

Table 4. Data Products

Pixel Data	
	- Imaging - overlapping FITS images
	- Atlas Images
	- Coadd image “stripe 82”
Catalog Data	
	- Legacy + SEGUE
	Photometric
	Spectroscopic
	- “Runs DB”
	includes overlaps/repeats
	Photometric only
	- “Stripe 82”
	Supernovae (time domain)
	Photometric only

Table 5. SDSS III.

Project	Description
BOSS	Dark Energy using baryon acoustic oscillations (upgraded spectrographs)
MARVELS	Planet Search (new spectrograph)
SEGUE-II	Continuation of SEGUE program
APOGEE	IR studies of Galactic structure (new spectrograph)

5. FUTURE

The SDSS is figuring strongly in the planning of current and future multiwavelength surveys. For example the UKIDSS sky coverage has been planned in large part to overlap with the SDSS survey as it existed three years ago. GALEX and Spitzer observations have also been planned, at least in part, to overlap with the SDSS footprint.

SDSS itself will continue in the future as SDSS-III. This time the programs are completely new (Table 5) and will include new institutions and have a new website.

Finally, the world is not standing still. Large new surveys such as PanStarrs, DES, LSST, VISTA, SPT, and LAMOST are in the planning or construction stages. The VO needs to be prepared to accommodate them.

ACKNOWLEDGMENTS

Funding for the SDSS and SDSS-II was provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS was managed by the Astrophysical Research Consortium for the Participating Institutions. Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy. This research has made use of NASA's Astrophysics Data System. This research has made use of data obtained from the US National Virtual Observatory, which is sponsored by the National Science Foundation.

REFERENCES

Gao, D., Zhang, Y.-X., and Zhao, Y.-H. 2008, MNRAS, 386, 1417

Santos, W. A., Mendes de Oliveira, C., and Sodré Jr., L. 2007, AJ, 134, 1551

Shuecker, P., Böhringer, H., and Voges, W. 2004, å, 420, 61

York, D. et al. 2000, AJ, 120, 1579