An Introduction to SDSS and its Database

Jurek Krzesinski[1]

1. Mt. Suhora Observatory, Pedagogical University of Cracow, ul. Podchorążych 2, 30-084 Cracow, Poland

The Sloan Digital Sky Survey is the largest and most successful sky survey of the Northern and part of the Southern celestial hemispheres ever made. The survey was performed using a dedicated 2.5 m telescope, the state-of-the-art imaging camera and spectrographs. It resulted in five band photometry of more than 14 000 square degrees of the sky, as well as over 4 000 000 unique spectra of galaxies, quasars and stars. The data are publicly available in the form of images, optical spectra, infrared spectra, and catalog data. All data can be downloaded from several different online sources. In this review I will present the past and current status of the Sloan Digital Sky Survey and wills introduce two main database web sites: *ScienceArchiveServer* and *SkyServer*.

1 Introduction

The idea of digital survey of the sky came in 1987 from James Edward Gunn who proposed building a dedicated CCD imager and spectrographs, putting them on a 2.5 m class telescope and scanning the visible sky. The project got the green light and became what we know today as the Sloan Digital Sky Survey (SDSS), from the name of the Sloan Foundation which covered most of the cost of the project. The construction site was chosen to be at Apache Point Observatory (APO) in the Sacramento Mountains of south-central New Mexico (latitude $32^\circ\ 46'\ 49.30''$ N, longitude $105^\circ\ 49'\ 13.50''$ W, elevation 2788 m). By the year 2000, the 2.5 m telescope (Fig. 1), with an extremely large 2.5 square degree field of view, the imager and the spectrographs (Gunn et al., 2006) were ready and the SDSS began routine operations.

Besides the Sloan Foundation, there were other US institutions participating in the project, as well as institutions from Japan and Germany. Since then more have joined the project, raising money and dedicating manpower[1].

In the beginning, the primary goal of the SDSS was to determine the structure of the Universe and quasars at large cosmic scales as well as the structure and kinematics of stars in our Galaxy. Today this part of the SDSS run is called the legacy survey (York et al., 2000). It was so successful that the SDSS was granted funds for subsequent years of operation and for new surveys: “The Sloan Extension for Galactic Understanding and Exploration” (SEGUE I and II) focused on the structure and evolution of our own galaxy, “The Baryon Oscillation Spectroscopic Survey” (BOSS) mapping the Universe on the largest scales, “The Multi-Object APO Radial Velocity Exoplanet Large-area Survey” (MARVELS) searching nearby stars for “exoplanets” and “The Apache Point Observatory Galactic Evolution Experiment” (APOGEE),

[1] see: `http://www.sdss.org/collaboration/` for the list of participants

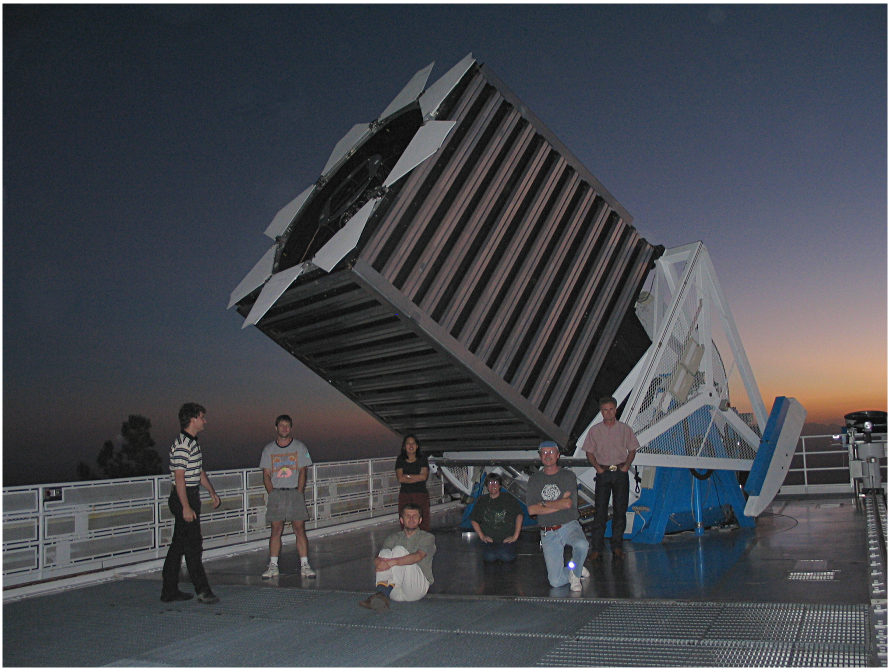

Fig. 1: The SDSS 2.5 m telescope and SDSS observers (J. Krzesinski 2002)

a high-resolution infrared spectroscopy survey focused on the structure and evolution of the Milky Way.

In January 2011, the SDSS-III released the "largest digital color image of the sky ever made" and, as stated by the BBC[2], it was published for professional astronomers and "citizen scientists" alike; in other words, it is free to anyone who wants a look. At the time of writing of this article (September 2015), the latest SDSS data release was in July 2014 (SDSS Data Release 12)[3]. In overview, the SDSS DR12 images were made in the Sloan u g r i z filters (Gunn et al., 1998; Doi et al., 2010) and cover 14 555 square degrees of the sky. DR12 contains data for 932 891 133 unique objects. The magnitude limits depend on the band in which a given object was observed and are up to $\sim$22 mag in u g r and $\sim$21 mag in i z filters. The effective area of the SEGUE and BOSS spectroscopy surveys covers over 10 693 square degrees of the sky. In total, the SDSS database contains 4 266 444 useful spectra of galaxies, quasars and stars. The magnitude limits of the spectroscopy depend on the sub-survey, signal-to-noise ratio and the object. For example, for galaxies the Petrosian r magnitude is less than $\sim$18 mag, for quasars the PSF i magnitude limit is less than $\sim$18 mag and the signal-to-noise ratio has to be greater than 4 per pixel at $g \approx 20$ mag). On top of these, the APOGEE infrared spectroscopic survey adds 156 593 unique spectra of stars from the galactic bulge, halo, disk and other fields.

If someone were to count the refereed papers which mention "SDSS" or "Sloan Digital Sky Survey" in the title or abstract, the number would be greater than 5000

[2] `http://www.bbc.com/news/science-environment-12167011`
[3] `http://www.sdss.org/dr12/scope/`

and these papers have been cited over 200 000 times.

2 Observing in SDSS

To understand some terms used in the SDSS database, for example: *run*, *camcol*, *field*, *plate* or *fiber*, one should know about the general construction of the instruments, i.e. the imaging camera and the multifiber spectrographs, and the way observations are or were performed in the SDSS. Let us start with the SDSS imaging camera[4] (Gunn et al., 1998). This used an array of 30 CCD chips arranged in six columns (*camcols* numbered from 1 to 6) to collect photometric images of the sky. Each CCD chip in a column was covered with SDSS *r i u z g* filters. The camera did not have any shutter, unlike most CCD cameras. Therefore, the SDSS telescope and thus the camera did not stay pointed at a single region of the sky taking a set of photometric images one by one, changing filters, opening and closing the shutter for exposures and then shifting the telescope to another field. Instead, to get an image, the telescope was slewing across the sky while in the focal plane images of the sky were drifting along the columns of CCD chips. As the images were crossing the CCDs, the charges accumulated in the rows of exposed CCD pixels were shifted in the direction of image drift and read at the drift rate. After crossing all 5 CCD chips in a given column the camera produced 5 images of the same region of the sky in *u g r i z* filters. The software pipeline processed images to produce so-called *fields* which consisted of 2048×1489 pixel frames of the same part of the sky in the 5 filters. Because it took a star 54 seconds to move from the beginning of a CCD to the end, this value was the effective exposure time in each filter. Because the CCD columns have some space between them, one slew of the telescope was not enough to make a complete scan of a particular strip of the sky. The strip had to be scanned at least a second time. Each scan was called a *run* and had its own unique number. Sometimes it took several nights in different months or years to complete a strip, therefore, strips can be composed of several *runs*.

When the images of the sky were ready, the objects for spectroscopic observations were software-selected from color-color cuts and assigned spectrograph optical *fibers*. The *fibers* pass the light from the focal plane to the SDSS spectrographs[5]. To hold the *fibers* in the telescope focal plane, a set of aluminium *plates* were made. Each *plate* had holes drilled corresponding to objects targeted for spectroscopic observation. To mount the *plates* in the focal plane, special cartridges with optical fibers were used. Every field in the sky (and object) has at least three spectroscopic observations to get the required signal-to-noise ratio and to get rid of cosmic rays. The *plates* and *fibers* also have their unique identification numbers, so for a specific date (expressed as the Modified Julian Date, MJD) we have a *plate* and a *fiber* number which correspond to a certain object in the sky.

The SDSS imager was used during photometric nights and replaced with cartridges for spectroscopic observations when the nights had bad seeing, clouds or too much moonlight. Most of the objects in the SDSS database have single *u g r i z* photometric observations and spectra, but due to overlaps of adjacent *runs*, *plates* or other observing programs, some of them have been observed twice or multiple times.

[4] `http://www.sdss.org/instruments/`
[5] `http://www.sdss.org/instruments/boss_spectrograph/`

3 SDSS database

The SDSS database has been developed since the beginning of the project and today we have excellent tools for viewing, searching and downloading the data. A natural consequence of the methods and equipment used for SDSS observations is searching the database by *run*, *camcol* and *field* numbers in the case of images, and by *MJD*, *plate* and *fiber* numbers for spectroscopy. However, this is more difficult and beginners can also start by searching for a given right ascension and declination or an object name.

There are two main interactive tools for searching the SDSS database: the Science Archive Server (SAS) for interactive spectra and image mosaics search, and the SkyServer giving access to the Catalog Archive Server (CAS). SAS[6] can be used to download calibrated images of the sky in $u\ g\ r\ i\ z$ filters and optical or infrared spectra of specified objects (if they exist). The data can be downloaded as FITS files (or JPEG in case of color digital photographs of the sky), therefore dedicated software to view these files is required. One of these, recommended by SDSS, is the `fv` package ("FITS viewer"). The SDSS interactive tools allow viewing of spectra and images directly using web browsers, but `fv` has wider capabilities of handling FITS files.

With the SkyServer[7] one can see a specified region in the sky (*navigate*), prepare a finding chart (*finding chart*) and query the SDSS database (*search*) for specific objects. The query restrictions can be made by spectroscopic or photometric parameters of the objects using dedicated search forms: *Rectangular*, *Search Form*, *Imaging Query* and *Spectro Query*. However, the most powerful query tool is the one based directly on the SQL language. Fortunately, this does not mean that one has to know the SQL language to start working with the SDSS database. The SkyServer has a large number of tutorials and examples of how to make SQL queries (see *Help* page and then *SQL Tutorial*). One can also prepare more sophisticated queries right in the query forms (*Search Form*) where an example of an SQL query can be generated right under the search form. One needs to remember that the SkyServer has some restrictions on the time a query can work. For the queries requiring a lot of computing time the best place to go is the *CasJobs*, a flexible advanced SQL-based interface to the CAS which requires (quick) registration. It can be accessed from the SkyServer *Data* → *casjobs* or directly[8].

These are not all the possibilities of the SkyServer, and many more options can be found at the links shown above. The scope of this article is only to help SDSS database users at the beginning of their work. For more information I recommend the descriptions, help and glossary available on the SDSS web pages. The most current SDSS data release and relevant links will be shown in any browser after typing just the word "SDSS" into the search box.

References

Doi, M., et al., AJ **139**, 1628 (2010)

Gunn, J., et al., AJ **116**, 3040 (1998)

[6] `http://dr12.sdss3.org/fields/runCamcolField?field=13&camcol=1&run=6122`
[7] `http://skyserver.sdss.org/dr12/en/home.aspx`
[8] `http://skyserver.sdss.org/casjobs/`

Gunn, J., et al., AJ **131**, 2332 (2006)
York, D., et al., AJ **120**, 1579 (2000)