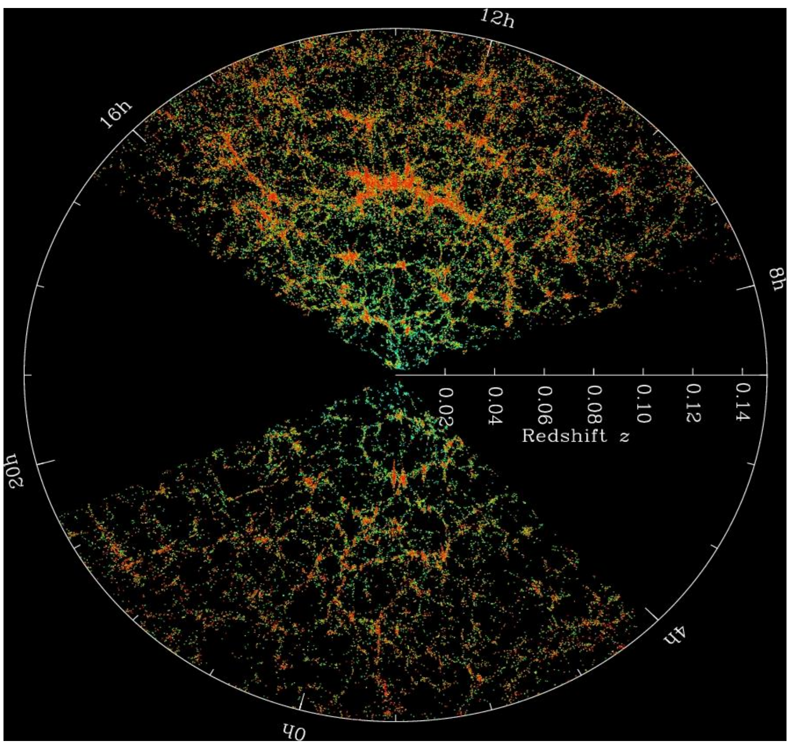
Studying galaxies with the Sloan Digital Sky Survey

Laboratory exercise, Physics of Galaxies, Spring 2019 (Uppsala Universitet)

by

Beatriz Villarroel



***

The Sloan Digital Sky Survey (SDSS) is the largest galaxy database existing today. Its imaging data covers one third of the sky and millions of objects have their spectra and photometry recorded in this database: stars, supernovae, galaxies. All these images were taken the last ten years at the Apache Observatory, using a 2.5 m wide angle telescope.

Fig. 1. The SDSS telescope building at the Apache Observatory, New Mexico, U.S.A.

The project has some very beautiful features for both professional astronomers as well as for amateurs, students and astronomy lovers – all the data is public and easily accessible for everybody. While many astronomical observations require access to big and powerful telescopes, or simulations so heavy that they only can be carried out on computer superclusters, any person possessing an ordinary desktop computer connected to the Internet, can do substantial research with the SDSS data.

It is our mission, to now explore some of the great potential of this database for galaxy research.

***

We start our journey through the Universe of the SDSS, by visiting the webpage of one of the later data release – the Data Release 12. (Although, in the exercises later we are going to use the more reliable and older Data Release 7.)

http://skyserver.sdss.org/dr12/en/home.aspx

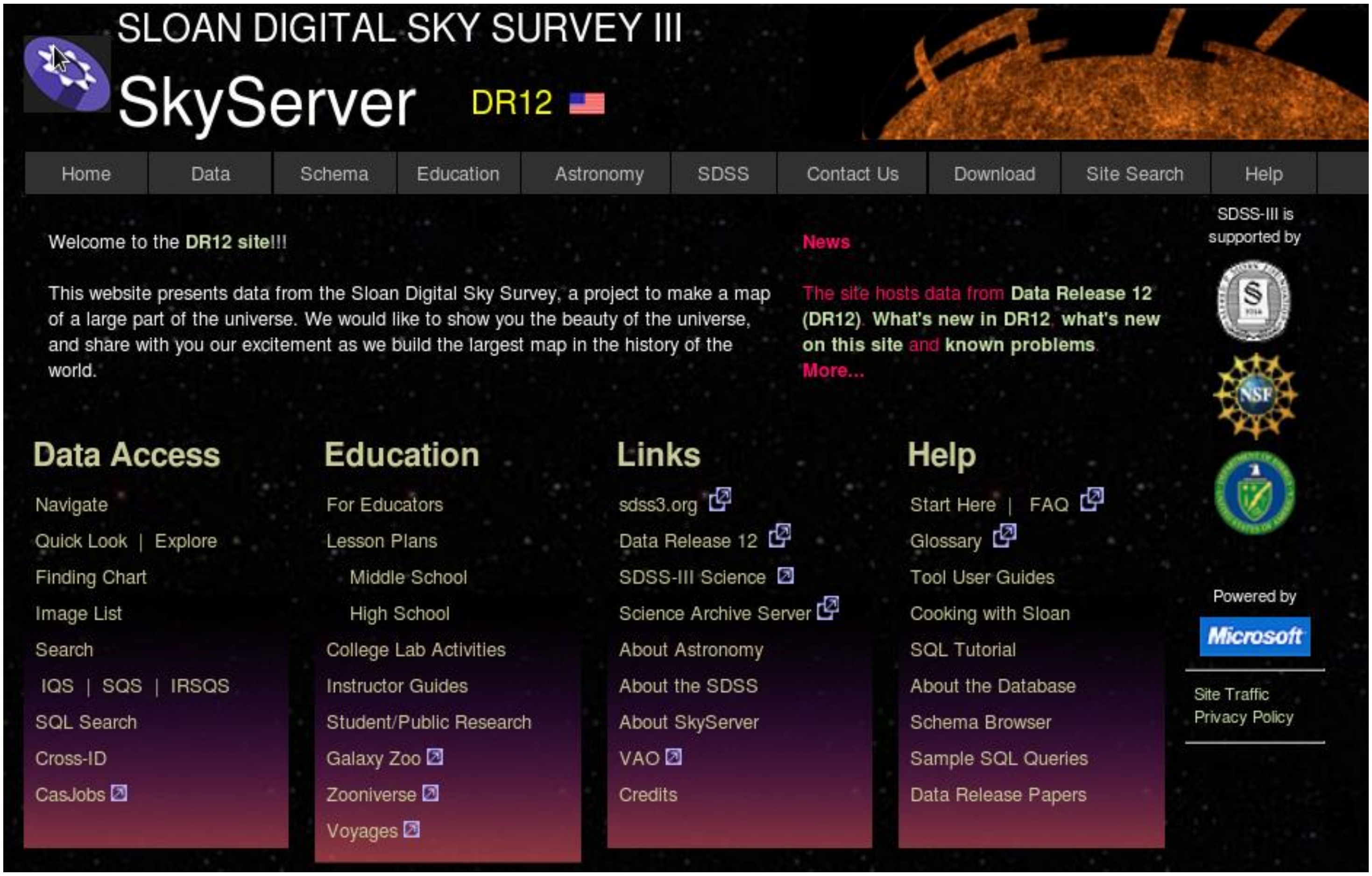


Fig. 2 The SDSS DR12 homepage.

This homepage has a rich collection of tools allowing you many different views on the same galaxies, objects or data. There are different types of data in the SDSS: images, spectra, photometric data and spectroscopic data. If you want to explore how the sky looks, you may use the Navigate Tool. The Casjobs page at the left bottom of the homepage is the one you will use the most to retrieve data.

***

Let us start the tour, by visiting the Famous Places of SDSS:

http://skyserver.sdss.org/dr12/en/tools/places/placeshome.aspx

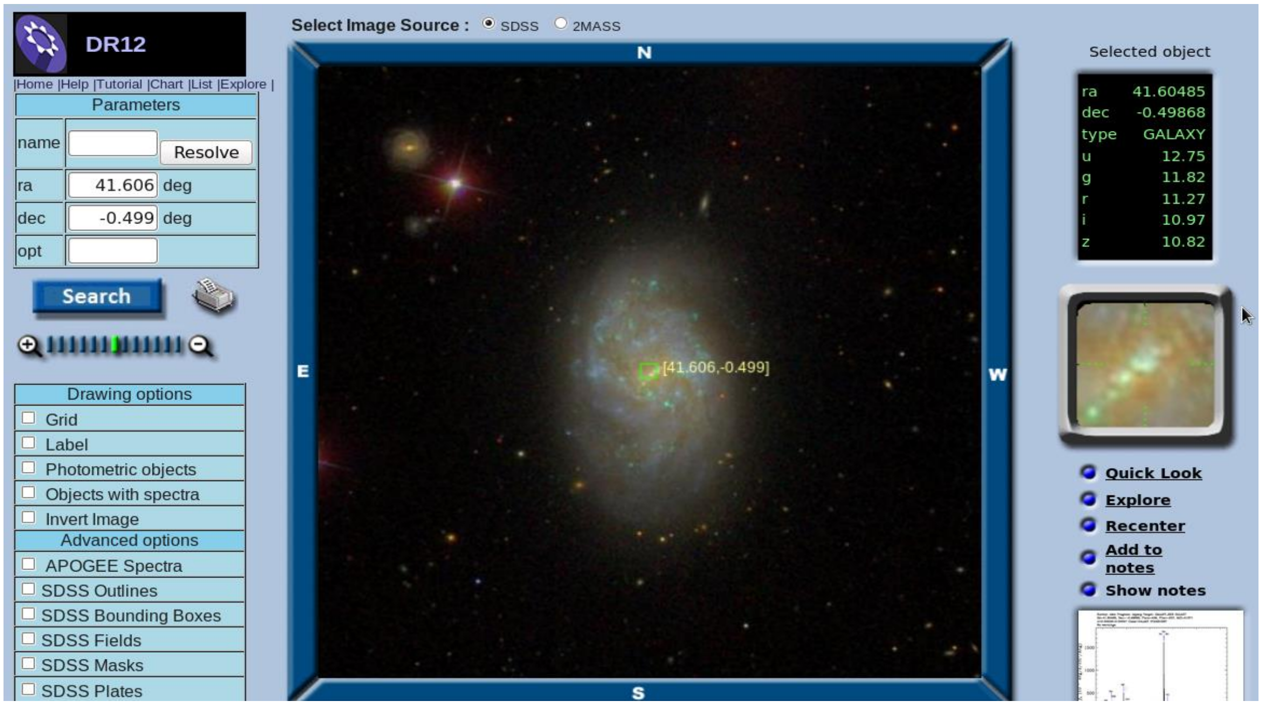


Fig.3. Center the green square at the object of interest. Try not to center it on a foreground star. The box in the upper, right corner shows the coordinates (right ascension, declination) in degrees. It also shows the galaxy type and the apparent magnitudes of the object in five different bands. In the lower right corner, is "Explore" that takes you to the page with additional information about the object and its optical spectrum.

Exercise 1: *Click on the image you find most beautiful or appealing. Click on the image and center the green square on the object of interest.*

a) *What type of object is it -- a star or a galaxy or something else? Note, that many foreground stars lie in front of galaxies and you want to make sure you select the right thing on the image.*

b) *What are the coordinates of the objects? What coordinate system are they specified by? (Hint: you might have to google for the coordinate system.)*

c) *Galaxies generally fall into two major populations: star-forming, blue ones and the red, passive galaxies. A lot of research effort goes into understanding how these two populations evolved. One way of distinguishing the colour of a galaxy, is by using the $u - r$ colour of the SDSS ugriz-system (Strateva et al. 2002). $u - r < 2.2$ counts as "blue", and all the other as "red". What $u - r$ colour does your object of interest have?*

d) *The $u - r$ colour often gets redder due to dust extinction. This also influences the spectrum so that some emission lines become weaker. What sources of dust can light emitted in a particular gas region in a distant galaxy experience on its way to us? Does the orientation of the observed galaxy matter?*

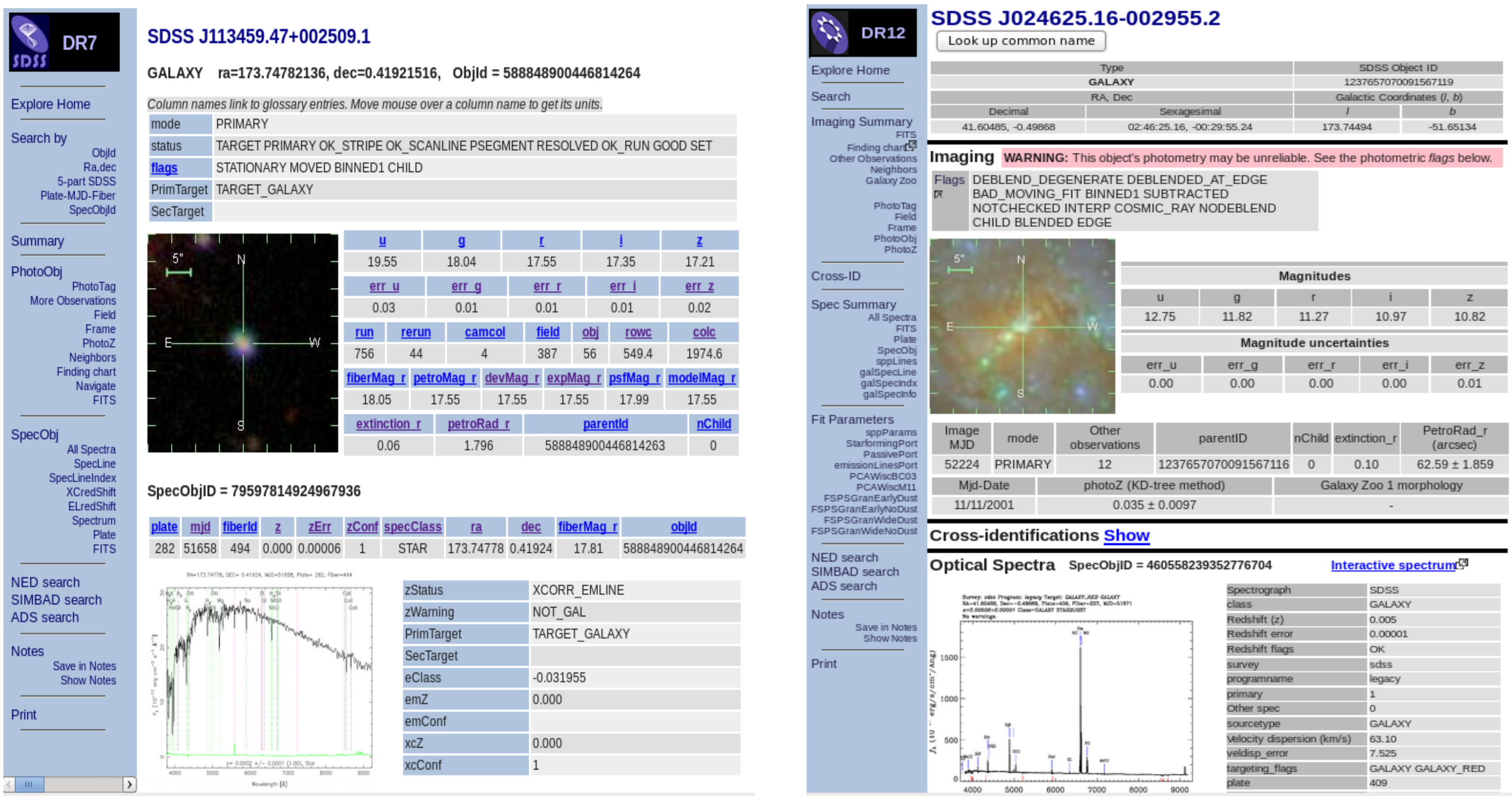


Fig. 4. The two different interfaces of the Explore Tool in Data Release 7 and Data Release 12. Beware: the object ID's have changed in between the two versions.

***

We continue the exercise by exploring the Explore Tool of the SDSS DR12:

http://skyserver.sdss.org/dr12/en/tools/explore/Summary.aspx?

...and the Explore Tool of the SDSS DR7:

http://skyserver.sdss.org/dr7/en/tools/explore/obj.asp

This is a very convenient tool you can use to look at the spectra, images and basic info for individual objects. The page shows you apparent magnitudes, redshifts, associated errors and sometimes also spectra of objects. You can find an object in different ways: you can use the typical SDSS name, the celestial coordinates or just the “object ID” (objID). The object ID is unique for every imaged object in the SDSS. In some catalogues it also is referred to as “bestObjID”.

Exercise 2:

Let us investigate three objects with the following IDs, using the old object explorer for the Seventh Data Release (DR7). Often, it is more advisable to use the DR7 catalogue as it has more raw measurements of the emission-lines and thus includes fewer biases introduced by model-dependent estimates.

We wish to compare the information in the Data Release 7 to the information in Data Release 12. But, oh no, in the DR12 the objects turn out to have new object IDs. Blasphemy! So now you have to find out a way to deduce the new name in the DR12 (hint: what observed quantity of a distant galaxy is unlikely to change within 100 years?)

object 1, DR7 objID:	*588010359067705540*
object 2, DR7 objID	*587738953104621690*
object 3, DR7 objID:	*588848900436525231*

a) *What are the new DR12 object ID's for each of the three objects?*

b) *Compare the apparent ugriz- magnitudes for each object in DR7 and DR12. What are the apparent magnitudes? What are the accompanying errors in their estimates? If there is a difference between the DR7 and DR12 apparent magnitudes, what is your conclusion about the error measurements?*

c) *What are the redshifts z for the objects? (Do not confuse them with the apparent magnitude in the z-band, also named z!)*

d) *Compare the three objects to each other. Describe in words what is similar or different in the images and spectra of the three objects.*

e) *Compare the three spectra to typical spectra for the following types of galaxies: Seyfert-1 galaxies, quasars, Seyfert-2 galaxies and star-forming galaxies. How do these classes of galaxies differ one from another? Make an advanced guess based on the spectra and motivate: which classes could either of these three objects belong to? Some typical spectra of active galactic nuclei can be found in Figure 5.*

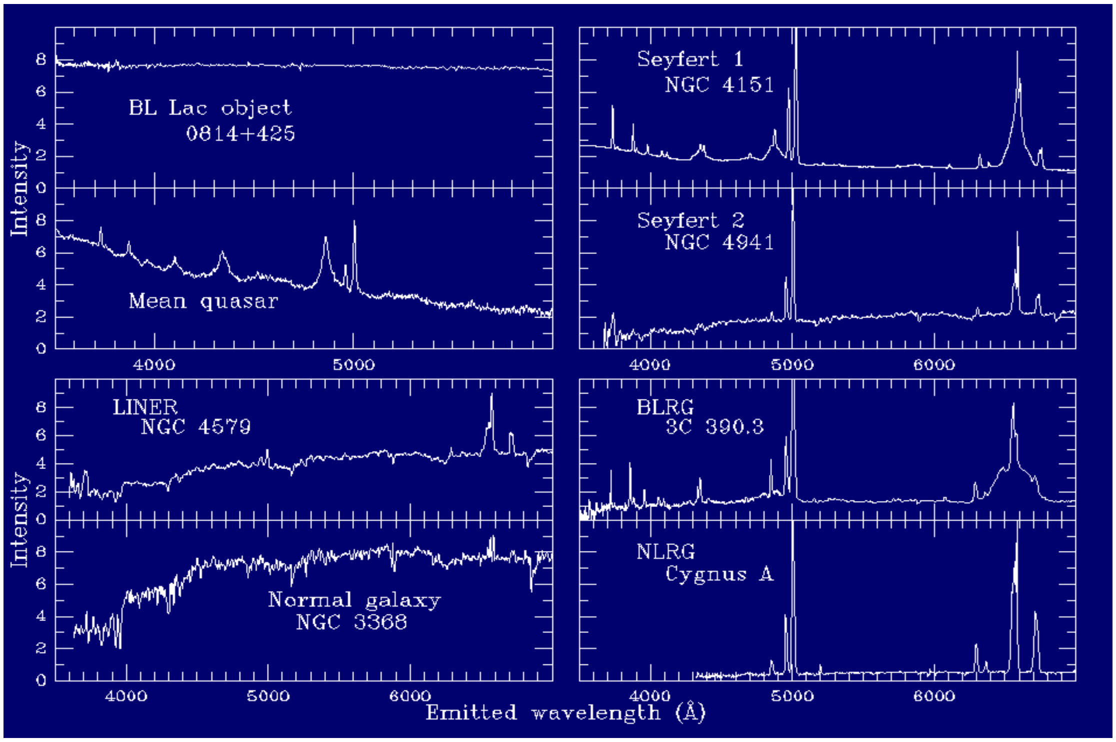


Fig. 5. Some typical galaxy spectra of active galactic nuclei and one normal, boring galaxy (NGC 3368).

***

Go to the SDSS start-page for the SDSS DR7.

http://skyserver.sdss.org/dr7/en/

Under **"Sky Server Tools"** (left column on the starting page), look for "**Casjobs**". You must register there one account. Please do so. This is the most convenient way for you to always find access back to your data, store your tables and delete whenever is needed. You will also get good feedback on all your errors *(oh yes, I expect you to have errors!)* when trying to submit your queries.

When inside your casjobs account, you can read the different sections:

a) **Query** is the place where you submit a query, designed to give you the desired sample.
b) **History**: everything you've ever done inside your account is kept track on. You can see the status of your queries.
c) **MyDB** is a databse with all the tables or samples that you have obtained so far.
d) **Import** allows you to upload your own tables to MyDB.
e) And whenever you decide to download a table from the MyDB to your own computer, the table (or sample) appears in "**Outpu**t".

***

The database has lots of different tables. Often, they overlap with other surveys e.g. ROSAT x-ray source catalogue, WISE IR-catalogue or the FIRST radio catalogue. To know what you are trying to download, you always have to go to the SDSS homepage. Do you see there, the word **Schema**? Click on it and you will soon notice the **Schema browser**. The Schema browser changes in between the data releases. Now you are going to need the one for the DR7:

http://skyserver.sdss.org/dr7/en/help/browser/browser.asp

The tables storing the objects you are interested in can be found under **Tables** and **Views**. Different functions to calculate different properties of the galaxies, can be found under **Functions** in the left column. Examples how the functions are used are given if you click on them.

This is all you are going to need. And a plotting program (Matlab?) for visualizing your data. Let us start the exercise.

Click on **Query**. Under **Context**, select "DR7". Whenever you submit a new query, write a name in **Table** and **Task Name**. Let the name for the table and the task name, be the same, always.

***

The Casjobs uses a SQL interface. The SQL language syntax is rather kind and does not care whether you use capitalized or non-capitalized letters. The code is built-up in the following manner: one section to specify what parameters you want to download (**select**), another section in what catalogues these parameters can be found (**from**) and a third section where logical conditions are introduced in order to specify your selection (**where**).

This is what builds up your query. The name of each catalogue used is convenient to shorten and the name you wish to give it is always specified after the real name, separated by a simple space. When you go to the query mode, you specify it under the yellow **Context** and select “DR7”, unless you already did what I told you to do.

We can identify various type of cosmic objects via the SpecObj catalogue that has lots of spectral information. The spectra are the essential ingredient in galaxy classifications.

In the SpecObj catalogue

specClass=2	stands for galaxies
specClass=3	quasars
specClass=4	high-redshift quasars

Now, if I'd like to download the redshifts and object id's (objid) of all quasars within redshift 0 < z < 1, I must write

select
S.bestobjID objectID, S.z redshift

from
SpecObj S

where
S.z > 0 AND S.z <1 AND (S.specclass=3 OR specClass=4)

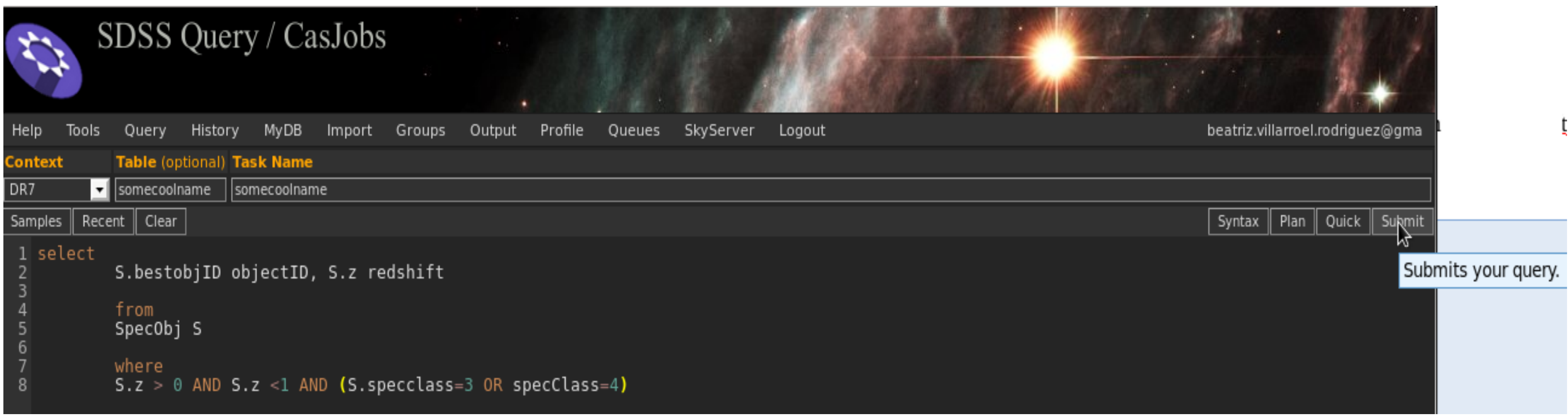


Fig.6. The Casjobs Query interface. Write your SQL code for selecting your favourite objects here.

Try this out. Clicking back in the "History" section, I can see the status of my submission. Difficult or data-heavy queries with many routines can take several hours, while an easy query takes just a few seconds or minutes. **Started** or **Finished** is good, while **Failed** means it didn't quite work out. Yet, the problem is often indicated in the **message** if you click on **jobID**. Many of the syntax errors can be noticed if you use the "Syntax" button before submitting the query.

***

Exercise 3:

Modify the above code and submit a query to get an estimate about the redshift distributions of all galaxies in the SDSS, and then of all objects defined as quasars. Go to ***MyDb****, click on the right table name in the left column. There, click on* ***Download,*** *chose "****Comma separated values****" (.csv-file). You will now be automatically redirected to a new window, called* ***Output****. There, click on the new* ***Download*** *and download the file.*

Use Matlab or any other suitable program to plot the histograms. Please note that software like Excel may have problems coping with data quantities this large. Include the plots in the report. In Matlab, the hist-function can be used by simply writing

hist(redshift)

where "redshift" is the list of all object redshifts from your sample you uploaded into Matlab.

a) *Observationally, what do the SDSS people mean by "quasars" or "galaxies"? What is your opinion on their nomenclature?*

b) *At which redshift do the two different populations peak in their number? What was the age of the Universe at this redshift?*

c) *Many observational biases play a role for what properties of objects we observe. Which biases do you imagine may exist in the observed redshift distributions?*

d) *What physics do you think plays a role in the above observed redshift distributions and the (eventual?) differences between the "galaxy" and the "quasar" populations? Can you think of any other factors that make the plots look like they do?*

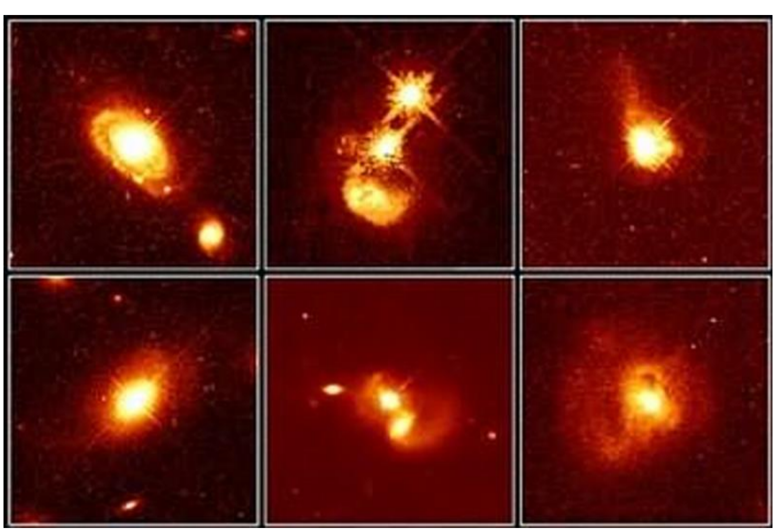


Fig. 7. Typical quasars.

***

Finally, I'd like you to reconnect this exercise to the **topic of your written essay or a specific research problem described in it**. Many research problems can be solved using the SDSS and sometimes one only has to figure out a clever way to approach them. Other times, even if one cannot solve the problem using the SDSS directly, the survey is helpful for e.g. selecting interesting candidate objects.

Exercise 4: *(a) Describe the most interesting research problem from the essay very briefly.*

(b) How could you approach this fascinating research problem using the SDSS? If it is not possible to use the SDSS, explain what the solution requires that cannot be met by the SDSS data.

(c) Finally, make a sketch of what information you need in a query from the SDSS to either solve that research problem or a part of it. Also describe how one should best use the obtained information/data. If your research problem cannot be solved with the SDSS, try to think of what steps towards a solution can benefit from this database.

A report containing answers to all exercises should be handed in individually. The report should ideally not be substantially longer than 5 pages. Take note that this exercise will be graded U, 3, 4, 5.

– THE END –