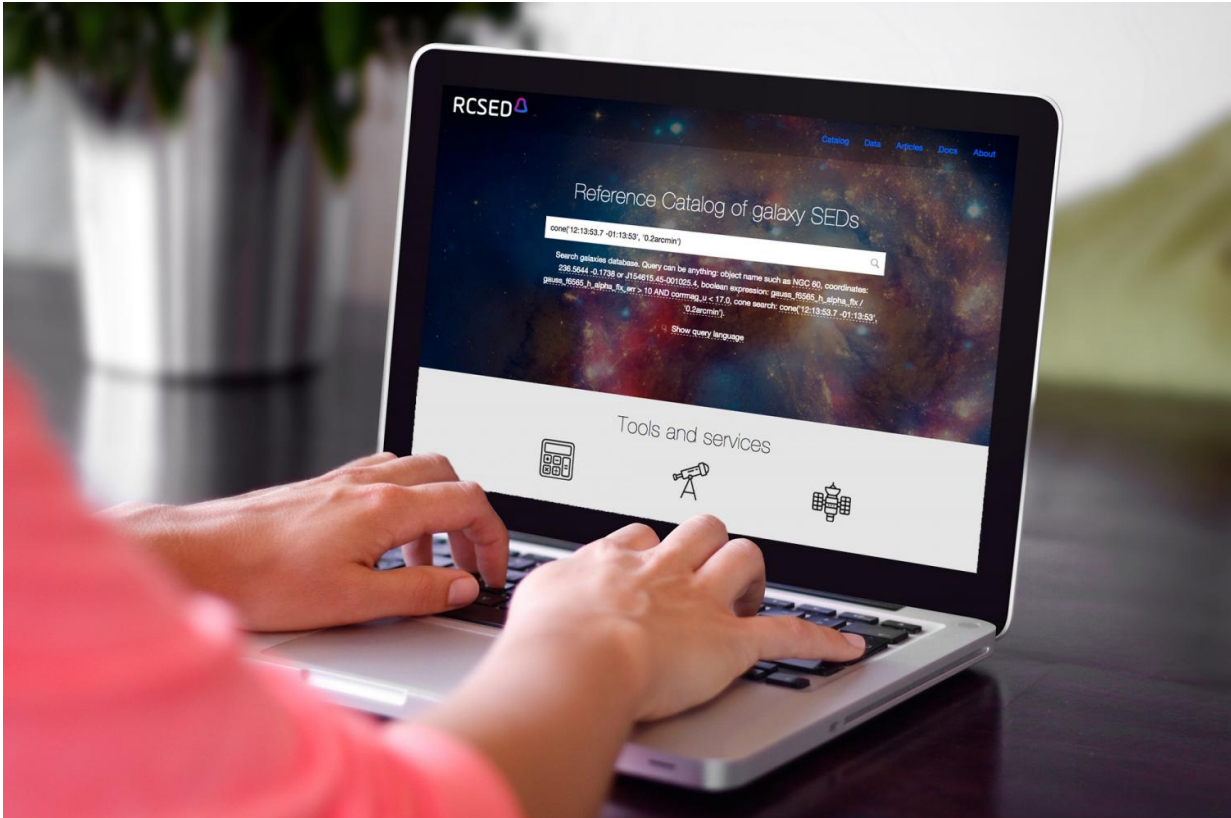# A Big Data approach to cataloging galaxies

February 9 2017

RCSED design. Credit: Ivan Zolotukhin

Astronomers at Lomonosov Moscow State University and collaborators have released "The Reference Catalog of galaxy SEDs" (RCSED), which contains value-added information about 800,000 galaxies. The catalog is accessible online, and the researchers have reported on their development in the *Astrophysical Journal Supplement*. Two co-authors

are undergraduate students at the Faculty of Physics, Lomonosov Moscow State University. While still working on the catalog, the team has published a few research papers based on their data, including a recent study in *Science*.

RCSED describes properties of 800,000 galaxies derived from elaborated data analysis. For every galaxy, it presents its stellar composition, brightness at ultraviolet, optical, and near-infrared wavelengths. From RCSED, researchers can also access galaxy spectra obtained by the Sloan Digital Sky Survey, measurements of spectral lines, and properties determined from this data, such as the chemical composition of stars and gas. This makes RCSED the first catalog of its kind that contains detailed homogeneous analysis for such a large number of objects.

Dr. Igor Chilingarian, an astronomer at Smithsonian Astrophysical Observatory, USA and a lead researcher at Sternberg Astronomical Institute, Lomonosov Moscow State University, says, "For every galaxy, we also provide a small cutout image from three sky surveys, which shows how the galaxy appears at different wavelengths. This provides us with the data for further investigations."

Dr. Ivan Katkov, a Senior Researcher at Sternberg Astronomical Institute adds, "The analysis of emission line profiles presented in RCSED is substantially more detailed and accurate then the data published in other catalogs."

RCSED is flexible and easy to use. By entering the object name or its coordinates in the search field, the website provides a single page of information referring to that object. Users can also access the catalog through Virtual Observatory applications such as TOPCAT. The RCSED website also provides tutorials including a technique that Igor Chilingarian and Ivan Zolotukhin exploited to discover new compact

elliptical galaxies, they published in the research paper "[Isolated compact elliptical galaxies: Stellar systems that ran away](link)."

Citizen scientists assisted in the development of the project website. Among them were high-level experts in software development and web design, who have daytime jobs in Russian tech companies.

Dr. Ivan Katkov adds: "The RCSED catalog was possible thanks to the application of an interdisciplinary Big Data approach, as we had to apply very complex scientific algorithms to a large dataset in a massively parallel way. Eventually, the expertise and resources available at large IT companies would undoubtedly allow researchers to significantly increase the quality and the quantity of research results and to make many important discoveries in astrophysics."

The fact that the RCSED catalog has attracted serious interest in the scientific community even during its assembly phase proves its great potential. During the last three years, several external researchers were given the access to the catalog on request and, using RCSED data, published over a dozen articles in peer-reviewed journals. The catalog is the world's largest homogeneous, value-added dataset for nearby galaxies, containing information collected with ground-based and space telescopes.

The current release of the RCSED catalog could have comprised a larger number of galaxies or contained extra bits of information about the currently included objects, but the scientists decided to focus on well-characterized datasets, which are described in detail and have known advantages and disadvantages. However, taking into account the project's importance for extragalactic astronomy and observational cosmology, the RCSED team is going to move forward and expand the catalog in the near future.

There are two principal directions of further RCSED development: the galaxy sample expansion and incorporating new data for existing objects. The team is considering including near- and mid-infrared data from the WISE satellite all-sky survey for the entire galaxy sample. However, this requires additional methodical work in order to homogenize the data for galaxies at different redshifts.

Moreover, it is possible to expand the principal galaxy sample by including spectra from the latest data release of the SDSS-III survey. This will turn 800,000 to 1.5 million objects.

Incorporating the publicly available spectral data from the Hectospec archive will add 300,000 to 400,000 objects at larger distances, whose spectra were collected with the 6.5-meter MMT telescope in Arizona. The current RCSED release comprises mostly nearby galaxies (by cosmological measures), whose redshifts are smaller than 0.4, because SDSS did not include faint objects. Therefore, the early universe is not represented in the catalog at all. The Hectospec archive will allow the team to move a little bit further along the cosmological distance scale until the redshift of 0.7. If they add several thousand galaxies from the DEEP2 survey conducted with the 10-meter Keck telescope in early 2000s, they could get insights into objects at redshift up-to 1.0, when the universe was less than half of its present age.

Igor Chilingarian concludes, "We shall be able to see the global picture in about 10 years, when large surveys like DESI have collected 25 to 30 million galaxy spectra out to intermediate redshifts."