Education and Heritage in the era of Big Data in Astronomy
Proceedings IAU Symposium No. 367, 2020
R. M. Ros, B. Garcia, S. R. Gullberg, J. Moldon & P. Rojo, eds.
doi:10.1017/S1743921321000570

The role of Big Data in Astronomy Education

A. M. Mickaelian and G. A. Mikayelyan

NAS RA V. Ambartsumian Byurakan Astrophysical Observatory (BAO),
Byurakan 0213, Aragatzotn Province, Armenia
email: aregmick@yahoo.com

Abstract. We review Big Data in Astronomy and its role in Astronomy Education. At present all-sky and large-area astronomical surveys and their catalogued data span over the whole range of electromagnetic spectrum, from gamma-ray to radio, as well as most important surveys giving optical images, proper motions, variability and spectroscopic data. Most important astronomical databases and archives are presented as well. They are powerful sources for many-sided efficient research using the Virtual Observatory (VO) environment. It is shown that using and analysis of Big Data accumulated in astronomy lead to many new discoveries. Using these data gives a significant advantage for Astronomy Education due to its attractiveness and due to big interest of young generation to computer science and technologies. The Computer Science itself benefits from data coming from the Universe and a new interdisciplinary science Astroinformatics has been created to manage these data.

Keywords. Big Data, Astronomical Surveys, Catalogues, Databases, Archives, Multiwavelength Astronomy, Multi-messenger Astronomy, Astroinformatics, Virtual Observatories.

1. Introduction

In Astronomy we deal with vast number of objects, phenomena and hence, big numbers. Astronomy and its results also enlarge most of other sciences, as any research on the Earth is limited in sense of the physical conditions, variety of objects and phenomena, and amount of data. During the last few decades astronomy became fully multiwavelength (MW); all-sky and large-area surveys and their catalogued data over the whole range of the electromagnetic spectrum from γ rays to radio wavelengths enriched and continue to enrich our knowledge about the Universe and supported the development of physics, geology, chemistry, biology and many other sciences. Astronomy has entered the Big Data era and these data are accumulated in astronomical catalogues, databases and archives. Astrophysical Virtual Observatories (VOs) have been created to build a research environment and to apply special standards and software systems to carry out more efficient research using all available databases and archives. VOs use available databases and current observing material as a collection of interoperating data archives and software tools to form a research environment in which complex research programs can be conducted. Most of the modern databases give at present VO access to the stored information. This makes possible not only the open access but also a fast analysis and managing of these data.

Present astronomical databases and archives contain billions of objects, both galactic and extragalactic, and the vast amount of data on them allows new studies and discoveries. Astronomers deal with big numbers and it is exactly the case that the expression

© The Author(s), 2021. Published by Cambridge University Press on behalf of International Astronomical Union

https://doi.org/10.1017/S1743921321000570 Published online by Cambridge University Press

"astronomical numbers" means "big numbers". Surveys are the main source for discovery of astronomical objects and accumulation of observational data for further analysis, interpretation, and achieving scientific results. Nowadays they are characterized by the numbers coming from the space; larger the sky and (in case of spectroscopic surveys) spectral coverage, better the spatial (in case of spectroscopic surveys, also spectral) resolution and sensitivity (deeper the survey), larger the covered time domain, more data are obtained and stored. Therefore, we give the highest importance to **all-sky and large area surveys**, as well as deep fields, where huge amount of information is available.

2. Multiwavelength era in astronomy

During many centuries optical wavelengths were the only source of information from the sky. However, modern astronomical research is impossible without various multiwavelength (MW) data present in numerous catalogues, archives, and databases. MW studies significantly changed our views on cosmic bodies and phenomena, giving an overall understanding and possibility to combine and/or compare data coming from various wavelength ranges. MW astronomy appeared during the last few decades and recent MW surveys (including those obtained with space telescopes) led to catalogues containing billions of objects along the whole electromagnetic spectrum. When combining MW data, one can learn much more due to variety of information related to the same object or area, as well as the Universe as a whole.

In Mickaelian (2016a) we list most important recent surveys (those having homogeneous data for a large number of sources over large area) and resulted catalogues providing photometric data along the whole wavelength range, from γ-ray to radio. All-sky and/or large area surveys have been carried out in many wavelengths covering a very wide range, from 300 GeV energies (or $4x10^{-18}$ A) to 74 MHz frequencies (or 4 m), which means a wavelength/frequency/energy ratio of 10^{-18}. Given that H.E.S.S. Gamma-ray telescope may observe up to 100 TeV energies (or 10^{-20} A) and LOFAR is designed for up to 10 MHz frequencies (or 30 m), this ratio reaches 10^{-21}. MW approach is applied in astrophysical research. Based on our estimate, MW astronomy provides 96 photometric points, out of which 64 come from all-sky or large area surveys, which means that these data are available for most of the studied sources, depending on the sensitivity.

3. Big Data Era in Astronomy

During the recent 2 decades, a number of giant projects were accomplished in astronomy completely changing the numbers of available information and requiring new approach in research. Among the biggest projects in astronomy one should mention the digitization of POSS I and II (**DSS I and II**) and creation of very big catalogues, **SDSS** with its accurate optical images and spectroscopy, **WISE** with very accurate positional and NIR/MIR photometric data that revolutionized astronomy in this wavelength domain, and **Gaia** with its unprecedently accurate astrometric data. Out of upcoming projects we would like to mention **LSST** and **SKA**. At present the biggest astronomical catalogs are the following: SuperCOSMOS (All-sky survey based on DSS1/2; 1,900,000,000 objects), Gaia EDR3 (All-sky; 1,811,709,771 objects), USNO B1.0 (All-sky; 1,045,913,669 objects), GSC 2.3.2 (All-sky; 945,592,683 objects), SDSS DR16 (covered area: 14,555 deg^2; the photometric catalog has 932,891,133 objects), AllWISE (NIR/MIR, All-sky; 747,634,026 objects), and 2MASS (NIR, All-sky; 470,992,970 objects).

Astronomical surveys give so much information that huge catalogues, dedicated archives and databases are being built to store, maintain and use these Big Data (Mickaelian 2016b). It is estimated that there are some 400 billion stars in the Milky Way galaxy and some 125-500 billion galaxies in the Universe, so that we are very far

https://doi.org/10.1017/S1743921321000570 Published online by Cambridge University Press

to catalogue all these objects. Even after Gaia space mission we will have much more accurate astrometric and photometric data for the stars but not much more completeness of detections. LSST and SKA will provide significantly more numbers, but again, full coverage of our estimated numbers in the Milky Way (stars) and especially in the Universe (galaxies and QSOs) will not happen in the nearest future.

Optical, UV and NIR/MIR wavelength ranges give most of the information from the sky, however MW astronomy was born in the recent decades and makes huge steps toward the overall understanding of the Universe with its various manifestations from γ-ray to radio and in the nearest future most of the objects (e.g. in our Galaxy or all galaxies in the Local Universe) will have their counterparts in all wavelengths. At present, approximate numbers of catalogued sources at different wavelength ranges are the following: γ-ray – 10,000, X-ray – 1,500,000, UV – 100,000,000, optical – 2,400,000,000, NIR – 600,000,000, MIR – 600,000,000, FIR – 500,000, sub-mm/mm – 200,000, radio – 2,000,000. These numbers give a comparative understanding about the wavelength coverage of the observed Universe. Of course, many of these sources represent the same astronomical object, however cross-correlations are not made for all these sources and we can only estimate the total number of detected astronomical objects, which may be some 3 billion.

As various astronomical missions, surveys, catalogues, databases and archives give various types of information, the only way to compare their sizes is to give this information in bytes. Astronomers, together with nuclear physicists, reach the largest possible numbers and put new requirements for computer science. As an example, LSST every night will provide 30 TB of data, which is much larger than many archives created and complemented during many years. The increase in data providing is happening due to covered sky areas and data accuracy, i. e. both resolution and sensitivity, as well as due to many times coverage, i. e. creation of possibilities for time domain studies.

4. Virtual Observatories

Astrophysical Virtual Observatories (VOs) have been created in a number of countries using their available databases and current observing material as a collection of interoperating data archives and software tools to form a research environment in which complex research programs can be conducted. The science goals are to define key requirements for large, complex MW astronomy projects. Interoperability includes the development and prototyping of new standards for data content, data description and data discovery. VO technology is the study and prototyping of Grid technologies that allow distributed computation, manipulation and visualization of data. A number of national projects have been developed in different countries since 2000, and an **International Virtual Observatory Alliance** (**IVOA**; www.ivoa.net) was created in 2002 to unify these national projects and coordinate the development of VO ideology and technologies. At present it involves 19 national and 2 European projects.

On IVOA webpage, there is a section for Education, "he VO for Students and the Public" (Tutorials, etc.) at https://ivoa.net/astronomers/vo_for_public.html. The European project EuroVO has developed special versions of professional astronomical tools which are adapted for use by school students, including improved performance on older PCs, and offline use. They have also developed a set of tutorials showing example usage of these tools. The tools and the tutorials can be found at http://vo-foreducation.oats.inaf.it//eng_download.html. It is worth mentioning that Google Sky uses the familiar Google Maps interface to display sky survey data together with various "showcase" objects from astronomical projects. A much richer interface is available in Google Earth tool, which enables users to add their own data.

https://doi.org/10.1017/S1743921321000570 Published online by Cambridge University Press

The Armenian Virtual Observatory (**ArVO**, www.aras.am/arvo.htm) was created based on the DFBS, Digitized Second Byurakan Survey (DSBS), and other digitization projects in Byurakan Astrophysical Observatory (BAO). ArVO project development includes the storage of the Armenian archives and telescope data, direct images and low-dispersion spectra cross-correlations, creation of a joint low-dispersion spectral database (DFBS / DSBS / HQS / HES / Case), a number of other science projects, etc. ArVO group at BAO was created in 2005 and it was authorized as an official project in IVOA also in 2005. An agreement on ArVO development between BAO and Institute for Informatics and Automation Problems (IIAP) was signed. The first science projects with DFBS/ArVO were the optical identifications of Spitzer Boötes sources in 2005. Joint projects were carried out between BAO and IIAP in 2007–2020. ArVO science projects are aimed at discoveries of new interesting objects searching definite types of low-dispersion spectra in the DFBS, by optical identifications of non-optical sources (X-ray, IR, radio) also using the DFBS and DSS/SDSS, by using cross-correlations of large catalogs and selection of objects by definite criteria, etc.

5. Summary

Modern astronomical data span the whole electromagnetic range (**multiwavelength astronomy**), as well as expand to neutrinos and gravitational waves and introduce **multi-messenger astronomy**. These are really Big Data with their volumes, variety of types and nature, velocity of generation and processing, and veracity of the quality (4 V-s). They provide the biggest data in all fields, as the Universe is much larger than the Earth and during the recent times astronomical technique became such that can observe, retrieve and analyze cosmic information in much better details than before.

Many of the modern **astronomical educational tools** use Big Data. This is because young people like computers, software and other computational tools, as well as because it is easier to make up online and automated tools for better distribution to the community. Using these data gives a significant advantage for Astronomy Education due to its attractiveness and due to big interest of young generation to computer science and technologies. The Computer Science itself benefits from Big Data coming from the Universe and a new interdisciplinary science Astroinformatics has been created to manage these data. And astronomical Virtual Observatories are the astronomical version of e-Science, which at present is being more and more developed.

References

Ahumada R., Allende Prieto C., Almeida A., et al. 2020, *ApJS 249, 3*

Gigoyan K. S., Mickaelian A. M., Kostandyan, G. R. 2019, *MNRAS 489, 2030. VizieR On-line Data Catalog III/266*

Lawrence A. 2007, *Astron. Geophys. 48, 3.27*

Markarian B. E., Lipovetsky V. A., Stepanian J. A., et al. 1989, *ComSAO 62, 5*

Massaro, E.; Mickaelian, A. M.; Nesci, R.; Weedman, D. (eds.) 2008, *The Digitized First Byurakan Survey, ARACNE Editrice, Rome, 78 p.*

Mickaelian A. M. 2016a, *Baltic Astron. 25, 75*

Mickaelian A. M. 2016b, *Astron. Rep. 60, 857*

Mickaelian A. M., Gigoyan K. S. 2006, *A&A 455, 765. VizieR On-line Data Catalog III/237A*

Mickaelian, A. M.; Nesci, R.; Rossi, C.; Weedman, D.; et al. 2007, *A&A, 464, 1177*

Véron-Cetty M.-P., Véron P. 2010, *A&A 518, 10*

https://doi.org/10.1017/S1743921321000570 Published online by Cambridge University Press