Big Data in Modern Astronomy and Computational Astrophysics

Areg Mickaelian

Byurakan Astrophysical Observatory (BAO)
Byurakan, Armenia
e-mail: aregmick@aras.am

ABSTRACT

The present state of all-sky and large area astronomical catalogues is given in γ-ray, X-ray, UV, optical, near/mid/far infrared, and radio, as well as most important surveys giving optical images, variability and spectroscopic data. Overall understanding of coverage along the whole wavelength range and comparisons between various surveys are discussed. Astronomy has entered the Big Data era. Astrophysical Virtual Observatories and Computational Astrophysics play an important role in using and analysis of big data for new discoveries.

Keywords

Astronomical surveys, Astronomical catalogues, Databases, Archives, Virtual Observatory, Big Data, Computational Astrophysics, Laboratory Astrophysics

1. INTRODUCTION

Recent large astronomical surveys both by ground-based and space telescopes and their catalogues during the last 15 years have accumulated vast amounts of data over the whole range of electromagnetic spectrum from Gamma-ray to radio. Present astronomical databases and archives contain billions of objects, both Galactic and extragalactic, and the vast amount of data on them allow new studies and discoveries. The Big Data era has arrived. Astrophysical Virtual Observatories (VO) 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. VO is a prototype of Grid technologies that allows distributed data computation, analysis and imaging. Particularly important are data reduction and analysis systems: spectral analysis, spectral energy distribution (SED) building and fitting, modeling, simulations, variability studies, cross-correlations, etc. The Numerical or Computational Astrophysics (part of Computer Science, also called Laboratory Astrophysics) has become an indissoluble part of astronomy and most of modern research is being done by means of it. Very often dozens of thousands of sources hide a few very interesting ones that are needed to be discovered by comparison of various physical characteristics. Cross-correlations result in revealing new objects and new samples. The large amount of data requires new approaches to data reduction, management and analysis. Powerful computer technologies are required, including clusters and grids. Large volume astronomical servers have been established to host Big Data and the International Council of Scientific Unions (ICSU) has recently created World Data System (WDS) to unify data coming from different science fields for further possibility of exchange and new science projects.

2. BIG DATA IN ASTRONOMY

2.1. Multiwavelength astronomical surveys and catalogues

Astronomy has benefited from multiwavelength (MW) studies, so that an overall understanding on any object is achieved. 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, such as e. g. the Milky Way (Fig 1).

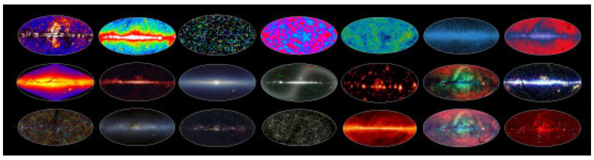

Figure 1. MW surveys imaging the entire sky and showing various patterns of the Milky Way.

Most important recent surveys and resulted catalogues are:

- **γ-ray.** Fermi-GLAST 3FGL catalogue (3,033 sources; Acero et al. 2015) and INTEGRAL (1,126 sources; Bird et al. 2010)
- **X-ray.** ROSAT BSC (18,806 sources; Voges et al. 1999) and FSC (105,924 sources; Voges et al. 2000), XMM (372,728 sources from various surveys), Chandra (380,000 sources from various surveys)
- **Ultraviolet (UV).** GALEX MIS (12,597,912 sources) and AIS (65,266,291 sources) (Bianchi et al. 2011)
- **Optical range.** SDSS DR12 (932,891,133 objects; Alam et al. 2015) and several POSS I and II based catalogues: APM (166,466,987 objects; McMahon et al. 2000), MAPS (89,234,404 objects; Cabanela et al. 2003), USNO B1.0 (1,045,913,669 objects; Monet et al. 2003, the largest catalogue so far by the number of objects), and GSC (945,592,683 objects; Lasker et al. 2008)
- **Near infrared (NIR).** 2MASS PSC (470,992,970 sources; Cutri et al. 2003) and ESC (1,647,599 sources; Skrutskie et al. 2006), DENIS (355,220,325 sources in Southern Sky; DENIS 2005)
- **Mid infrared (MIR).** WISE (563,921,584 sources; Cutri et al. 2013), IRAS PSC (245,889 sources; IRAS 1988) and FSC (173,044 sources; Moshir et al. 1990) (IRAS joint catalogue has been compiled containing 345,162 sources, Abrahamyan et al. 2015), AKARI

IRC (870,973 sources; Ishihara et al. 2010), SST (4,261,028 sources from various surveys)

- **Far infrared (FIR).** IRAS (345,162 sources, Abrahamyan et al. 2015), AKARI FIS (427,071 sources; Yamamura et al. 2010)
- **Radio.** NVSS (1,773,484 sources at 1.4 GHz; Condon et al. 1998), FIRST (946,432 sources at 1.4 GHz; Helfand et al. 2015), SUMSS (211,063 sources at 843 MHz; Mauch et al. 2003), WENSS (229,420 sources at 330 MHz; de Bruyn et al. 1998), GB6 (75,162 sources at 4.83 GHz; Gregory et al. 1996), 7C (43,683 sources at 151 MHz; Hales et al. 2007)

Summarizing, astronomers deal with the following numbers in various wavelength ranges: 10,000 sources in γ-ray, 1,500,000 sources in X-ray, 100,000,000 sources in UV, 1,500,000,000 objects in optical range, 600,000,000 sources in NIR, 600,000,000 sources in MIR, 500,000 sources in FIR and 2,000,000 sources in radio.

2.2. Optical images, variability and spectra

Beside the main MW catalogues giving photometric data, there have been a number of astronomical surveys aimed at covering large areas and obtaining optical images, measuring variability and obtaining spectroscopic data. SDSS gives both photometric and spectral data in large area (Fig. 2).

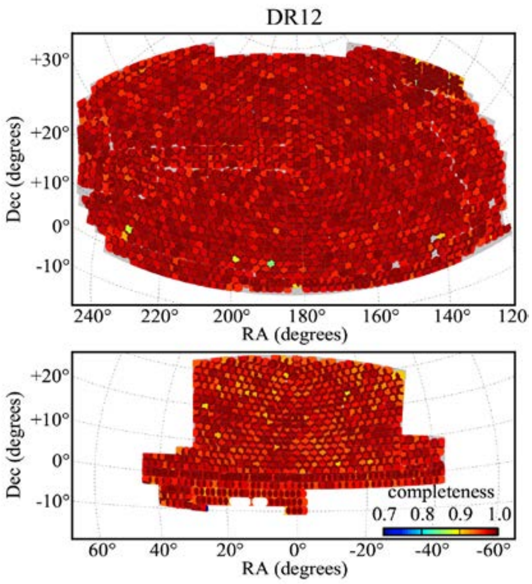

Figure 2. SDSS sky coverage.

Most important among these surveys and catalogues are:

- **Optical images.** Palomar Observatory Sky Surveys (POSS) I and II and their digitized versions DSS I (McGlynn et al. 1994) and DSS II (Lasker et al. 1996) cover the entire sky. Most of objects are catalogued in USNO B1.0 and GSC. SDSS covers 14,555 deg^2 and is the most accurate large survey (Alam et al. 2015)
- **Variability.** Variable objects are being obtained by repeated observations and are given in a number of catalogues: GCVS (80,671 variables; Samus et al. 2011), All Sky Automated Survey (ASAS) Catalogues of Variable Stars (ACVS, 30,000 new objects; Pojmanski 1998), NSVS (14,000,000 objects, including 8678 red variables; Wozniak et al. 2004), Catalina Sky Survey (CSS) Cataclysmic Variables (Drake et al. 2014a), Periodic Variable Stars (Drake et al. 2014b) and RR Lyrae stars (Abbas et al. 2014), and Pan-STARRS. Most complete resource is the International Variable Star Index, 203,438 objects (www.aavso.org/vsx). An understanding on the variability of large amount of objects may be obtained by comparing POSS I and POSS II photometric data, as shown by Mickaelian et al. (2011), as well as the same method is valid for proper motions (Mickaelian & Sinamyan 2010)
- **Spectroscopy.** Large amount of spectra have come from objective prism surveys, such as FBS (20,000,000 low-dispersion spectra; Markarian et al. 1989), FBS (Markarian et al. 1989), SBS (Stepanian 2005), Case (Pesch et al. 1995), HQS (Hagen et al. 1999) and HES (Wisotzki et al. 2000). FBS (Mickaelian et al. 2007) and HQS have been digitized and are available online. Most important modern spectroscopic surveys (typically galaxy redshift or QSO surveys) are 2dF/6dF (346,061 galaxies and 49,425 stellar objects, including 23,660 QSOs; Colless et al. 2001, Croom et al. 2004) and SDSS (4,355,200 spectra; including 2,401,952 galaxies, 477,161 QSOs and 851,968 stars; Alam et al. 2015). CALIFA is mapping 600 galaxies with imaging spectroscopy (IFS) and produces more than 1 million spectra. GAMA spectroscopic survey is for 300,000 galaxies also providing millions of spectra

Summarizing, we have many times covered the entire sky with optical imaging, as well as have obtained deeper images for selected smaller areas (deep fields). More than 200,000 variable objects have been discovered. At present some 5 million objects (compared to 300,000 ones 20 years ago) have spectroscopy giving understanding on their nature and possibility of detailed investigation. The number of QSOs doubles every 5 years (Véron-Cetty & Véron 2010; Paris et al. 2014). Thousands of Blazars have been identified (Massaro et al. 2015).

3. ASTRONOMICAL DATABASES AND ARCHIVES

All data obtained from astronomical observations are being archived for further use. Astronomical databases are being built for efficient usage of the accumulated data.

- **Wide Field Plate Data Base** (**WFPDB**; http://www.skyarchive.org; Tsvetkov et al. 1994) contains information on all wide-field (>1°) photographic observations. The total number of plates obtained during 1879-2002 in 125 observatories with more than 200 telescopes is 2,204,725 collected from 345 archives. 2,128,330 are direct images and only 64,095 are objective prism plates, including 2500 FBS and SBS plates
- Among the most important **modern archives** are ESO Archive (archive.eso.org/eso/eso_archive_main.html) and multi-mission archives and servers maintaining data from space telescopes (High Energy Astrophysics Science Archive Research Center, HEASARC, http://heasarc.gsfc.nasa.gov; NASA/IPAC Infra-Red Science Archive, IRSA, http://irsa.ipac.caltech.edu; Multimission Archive at Space Telescope, MAST; http://archive.stsci.edu and others)
- **Centre de Données Astronomiques de Strasbourg** (**CDS**; http://cdsweb.u-strasbg.fr/) was created in 1972 and is being updated permanently. CDS contains all possible information on astronomical objects. It has several important services: **SIMBAD** (Set of Identifications, Measurements, and Bibliography for Astronomical Data; http://simbad.u-strasbg.fr/simbad/)

– world largest database of astronomical objects, **VizieR** (http://vizier.u-strasbg.fr/viz-bin/VizieR; Ochsenbein et al. 2000) – database of astronomical catalogues and published tables containing some 14,000 catalogues and 12,000 tables), **Aladin** (http://aladin.u-strasbg.fr/aladin.gml) – an interactive sky atlas allowing overplot objects from various catalogues and study together in the same field

- Large extragalactic databases: **NED** (NASA/IPAC Extragalactic Database; http://ned.ipac.caltech.edu/) containing data on more than 168 million objects and **HyperLEDA** (http://leda.univ-lyon1.fr/) containing detailed data for study of physics and evolution of galaxies
- Other databases contain data on astronomers, astronomical literature, preprints, abstracts, etc. Most important are **ADS** (SAO/NASA Astrophysics Data System; adsabs.harvard.edu/abstract_service.html, a digital library portal maintaining more than 9.4 million entries on astronomy/astrophysics, physics and arXiv e-prints) and **astro-ph** (arxiv.org/archive/astro-ph, maintains astronomical preprints since 1992 on physics, mathematics, informatics, and statistics)

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 18 national and 2 European projects.

IVOA has Working Groups on Semantics, Data Access Layer, VO Event, Data Modeling, Resource Registry, Grid & Web Services, and VOTable and Interest Groups on Theory, Open Grid Forum Astronomy Research Group (OGF Astro-RG), Data Curation & Preservation, Knowledge Discovery in Databases. IVOA software and tools relate to Data discovery (Aladin, Astroscope, VOExplorer, Datascope), Spectral analysis (VOSpec, SPLAT, EURO-3D, Specview), Data visualization and handling (VOPlot, Topcat, VisIVO, STILTS), Spectral Energy Distribution (SED) building and fitting (VOSED, Yafit, easy-z, GOSSIP), etc. Spectral analysis tools allow combining spectral data coming from various telescopes at different wavelengths and joint analysis for line measurements, matching with theoretical models, etc., as for example in VOSpec. Building SEDs for AGN allow having an overall understanding on their energy distribution and better classifications.

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 Bootes sources in 2005. Joint projects were carried out between BAO and IIAP in 2007-2010 (Mickaelian et al. 2009a; 2009b). VO lectures are being regularly given at the Byurakan International Summer Schools (2006, 2008, 2010 and 2012).

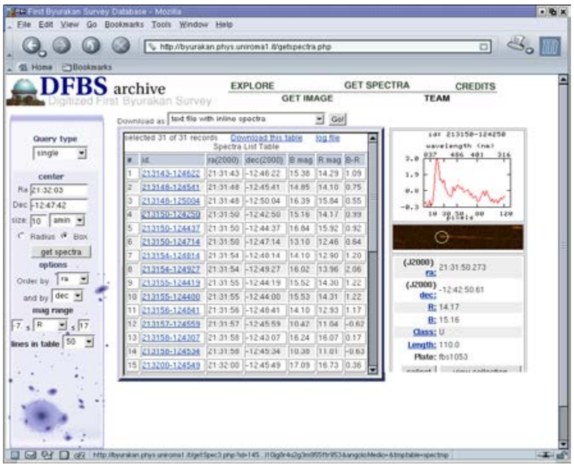

Figure 3. *DFBS portal in Get Spectra mode.*

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. COMPUTATIONAL ASTROPHYSICS

Modern astronomical research is impossible without various MW data present in numerous catalogues, archives, and databases. A user is able to search for any data in them, cross-correlate and make a comparative analysis. Surveys are much more valuable when various data can be compared and studied together. That is why it is so important to have easy access to all databases in a standard way. This is the task of the VOs. A number of efficient research projects have become possible, such as data discovery, spectral reduction and analysis, image processing, SED building and fitting, modeling, simulations, variability studies, cross-matching (cross-correlations), etc.

Dedicated astronomical software is especially important to achieve the needed tasks. The main standard of astronomical data is **FITS (Flexible Image Transfer System)**. It is being used in most of the software and systems. Most important software systems are **MIDAS (Munich Image Data Analysis System)** and **IRAF (Image Reduction and Analysis Facility)**. ESO-MIDAS system (www.eso.org/sci/software/esomidas/) provides general software for image and spectra reduction and analysis, measurements, calculations, building graphs, work with tables, etc. IRAF is a similar universal system (iraf.noao.edu/). IRAF commands are based on package structures, so that additional packages may be added. Examples of IRAF applications are calibrations of fluxes and positions of astronomical objects, correction of variations of sensitivity between pixels of the receiver, matching multiple images or measuring redshifts of absorption and emission lines in the spectrum.

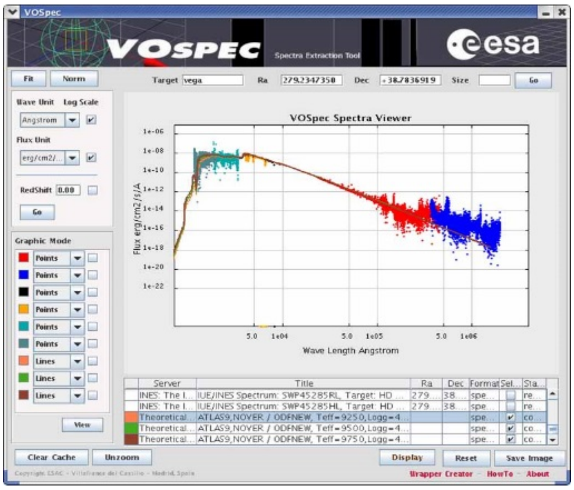

Figure 4. *VOSpec spectral analysis software showing how different spectral data may be combined and analyzed.*

In Astrophysics, main results of the 20th century related to accomplishment of theoretical problems by using analytical methods. However, the complexity of many astrophysical phenomena shows that analytical methods are available only for limited cases. Therefore, to understand astrophysical phenomena, numerical methods have become irreplaceable and promise to have dominant role in the methodology of theoreticians. Very important is the presence of Big Data, which is the fourth axis of modern science (Hey et al. 2009). At present it is impossible to separate high performance computations and big data, as there is a need to analyze the vast amount of data coming from various telescopes, large instruments, space facilities and other sources.

The **Computational Astrophysics** (or **Laboratory Astrophysics** is also used) has become an important part of astronomical research, without which modern results are impossible.

REFERENCES

[1] Abbas M. A., Grebel E. K., Martin N. F., et al. 2014, MNRAS 441, 1230
[2] Abrahamyan, H. V.; Mickaelian, A. M.; Knyazyan, A. V. 2015, A&C 10, 99
[3] Acero, F.; Ackermann, M.; et al. 2015, ApJS 218, 23
[4] Alam, S.; Albareti, F. D.; et al. 2015, ApJS 219, 12
[5] Bianchi, L.; Herald, J.; Efremova, B.; et al. 2011, Ap&SS 335, 161
[6] Bird, A. J.; Bazzano, A.; Bassani, L.; et al. 2010, ApJS 186, 1
[7] Cabanela, J. E., Humphreys, R. M., Aldering, G., et al.: 2003 PASP 115, 837
[8] Colless, M.; Dalton, G.; Maddox, S.; 2001, MNRAS 328, 1039
[9] Condon, J. J., Cotton, W. D., et al.: 1998 AJ 115, 1693
[10] Croom, S. M.; Smith, R. J.; Boyle, B. J.; et al. 2004, MNRAS 349, 1397
[11] Cutri, R. M.; Skrutskie, M. F.; van Dyk, S.; et al. 2003, IPAC/California Institute of Technology
[12] Cutri, R. M., et al. 2013, IPAC/Caltech
[13] DENIS consortium 2005, Vizier online data catalogue
[14] de Bruyn G., Miley G., Rengelink R., et al. 1998, WENSS Collab. NFRA/ASTRON and Leiden Obs.
[15] Drake A. J., Gansicke B. T., Djorgovski S. G., et al. 2014a, MNRAS 441, 1186
[16] Drake A. J., Graham M. J., Djorgovski S. G., et al. 2014b, ApJS 213, 9
[17] Gregory, P. C., Scott, W. K., Douglas, K., Condon, J. J. 1996, ApJS 103, 427
[18] Hagen, H.-J.; Engels, D.; Reimers, D. 1999, A&AS 134, 483
[19] Hales, S. E. G.; Riley, J. M.; Waldram, E. M.; et al. 2007, MNRAS 382, 1639
[20] Helfand, D. J.; White, R. L.; Becker, R. H. 2015, ApJ 801, 26
[21] Hey, T.; Tansley, S.; Tolle, K. (Eds.) 2009, The Fourth Paradigm: Data-Intensive Scientific Discovery
[22] IRAS, 1988, Joint IRAS Scienve Working Group. IRAS PSC, Version 2.0, NASA RP-1190
[23] Ishihara, D.; Onaka, T.; et al. 2010, A&A 514, 1
[24] Lasker, B. M., Doggett, J., McLean, B., et al. 1996, ASP Conf. Ser. 101, 88
[25] Lasker, B. M., Lattanzi, M. G., McLean, B. J., et al. 2008, AJ, 136, 735L
[26] Markarian, B. E.; Lipovetsky, V. A.; Stepanian, J. A.; et al. 1989, Comm. SAO 62, 5
[27] Massaro, E.; Maselli, A.; et al. 2015, Ap&SS 357, 75
[28] Mauch, T.; Murphy, T.; Buttery, H. J.; et al. 2003, MNRAS 342, 1117
[29] McGlynn, T., White, N. E., Scollick, K., 1994, ASP Conf. Ser., 61, 34
[30] McMahon, R.G., Irwin, M.J., Maddox, S.J. 2000, The APM-North Catalogue, Inst. of Astron., Cambridge, UK
[31] Mickaelian, A. M.; Nesci, R.; Rossi, C.; et al. 2007, A&A 464, 1177
[32] Mickaelian A. M., Sargsyan L. A., Astsatryan H. V., et al. 2009a, Data Science Journal 8, 152
[33] Mickaelian A. M., Astsatryan H. V., Sahakyan V. G., et al. 2009b, Proc. CSIT-2009 meeting, p. 420
[34] Mickaelian A.M., Mikayelyan G.A., Sinamyan P.K. 2011, MNRAS 415, 1061
[35] Mickaelian A., Sinamyan P. 2010, MNRAS 407, 681
[36] Monet, D. G.; Levine, S. E.; et al. 2003, AJ 125, 984
[37] Moshir, M., Kopan, G., Conrow, T., et al. 1990, IRAS FSC, Version 2.0, NASA
[38] Paris I., Petitjean P., et al. 2014, A&A 563, 54
[39] Pesch, P.; Stephenson, C. B.; MacConnell, D. J. 1995, ApJS 98, 41
[40] Pojmanski G. 1998, Acta Astron. 48, 35
[41] Samus', N. N., Durlevich, O. V., Zharova, A. V., et al. 2011, GCVS database, Inst. Astron. and SAI, Moscow
[42] Skrutskie, M. F.; Cutri, R. M.; et al. 2006, AJ 131, 1163
[43] Stepanian, J.A. 2005, RMxAA 41, 155
[44] Tsvetkov, M.K., Stavrev, K.Y., Tsvetkova, K.P., et al. 1994, Proc. IAU S161, Kluwer, Dordrecht, p. 359
[45] Véron-Cetty, M.-P.; Véron, P. 2010, A&A 518, 10
[46] Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389
[47] Voges, W., Aschenbach, B., Boller, T., et al. 2000, MPE Garching
[48] Wisotzki, L.; Christlieb, N.; Bade, N.; et al. 2000, A&A 358, 77
[49] Wozniak, P. R., Vestrand, W. T., Akerlof, C. W., et al. 2004, AJ 127, 2436
[50] Yamamura, I.; Makiuti, S.; Ikeda, N.; et al. 2010, AKARI/FIS All-Sky Survey Point Source Catalogues