Communications of BAO, Vol. 67, Issue 2, 2020, pp. 159-180

Big Data in Astronomy: Surveys, Catalogs, Databases and Archives

A. M. Mickaelian*

NAS RA Byurakan Astrophysical Observatory (BAO), Byurakan 0213, Aragatzotn Province, Armenia

Abstract

We present the modern situation in astronomy, where Big Data coming from the Universe put new tasks for catalogizing, storage, archiving, analysis and usage of the scientific information. The two major characteristics of modern astronomy are multiwavelength (MW) studies (from γ-ray to radio, as well as multi-messenger studies, using also neutrinos, gravitational waves, etc.) and Big Data (including data acquisition, storage and analysis). Present astronomical databases and archives contain billions of objects observed in various wavelengths, both Galactic and extragalactic, and the vast amount of data on them allows new studies and discoveries. Astronomers deal with 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. We review the main characteristics of astronomical surveys, we compare photographic and digital eras of astronomical studies (including the development of wide-field observations), we give the present state of MW surveys, and we discuss the Big Data in astronomy and related topics of Virtual Observatories and Computational Astrophysics. The review includes many numbers and data that can be compared to have a possibly overall understanding on the studied Universe, cosmic numbers and their relationship to modern computational possibilities.

Keywords: *Big Data, Astronomical Surveys, Astronomical Catalogues, Databases, Archives, Multiwavelength Astronomy, Data Mining, Computational Astrophysics, Astrostatistics, Astroinformatics, Virtual Observatories, Laboratory Astrophysics.*

1. Introduction

Astronomy is the area of science where 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. 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

*aregmick@yahoo.com

DOI: 10.52526/25792776-2020.67.2-159

fitting, modelling, variability studies, cross-correlations, etc. This way, astronomers benefit from the usage of data coming from various ground-based and space telescopes, from various observing modes, from various time domains and from various wavelength ranges. Putting together all data (including old archival ones) allows discovering new objects, studying variability and finding high proper motion stars. Therefore, general knowledge of the important astronomical surveys, catalogues, databases and archives and main parameters of the available data is necessary for modern astrophysical research. Moreover, these accumulated data and the necessity of their quick reduction and analysis, as well as modelling and simulations in theoretical studies led to the creation of the Numerical or Computational Astrophysics (Astrostatistics, Astroinformatics), which is part of the Computer Science. It has become an indissoluble part of astronomy and most of modern research is being done by means of it. On the other side, Laboratory Astrophysics provides laboratory experiments related to space research to check the results by astronomical observations.

Big Data are characterized by 4 Vs:

- **Volume.** Quantity of generated and stored data. The size of the data determines the value and potential insight, whether it can be considered big data or not.
- **Variety.** The type and nature of the data. This helps people who analyze it to effectively use the resulting insight. Big data draws from text, images, audio, video; plus, it completes missing pieces through data fusion.
- **Velocity.** In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Big data is often available in real-time.
- **Veracity.** The data quality of captured data can vary greatly, affecting the accurate analysis.

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 "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. These are:

- CGRO (Hartman et al., 1999), Fermi-GLAST (Acero et al., 2015) and INTEGRAL (Bird et al., 2010) in **γ-ray**,
- ROSAT (Voges et al., 1999, 2000), Swift (D'Elia et al., 2013), XMM-Newton (XMM-Newton, 2013) and Chandra (Evans et al., 2010) in **X-ray**,
- GALEX (Bianchi et al., 2011) in **UV**,
- SDSS (Ahumada et al., 2020) and POSS I / POSS II based several catalogues (APM (McMahon et al., 2000), MAPS (Cabanela et al., 2003), USNO (Monet et al., 1998, 2003) and GSC (Lasker et al., 2008)) in **optical range**,
- 2MASS (Cutri et al., 2003, Skrutskie et al., 2006) and DENIS (DENIS consortium, 2005) in **near infrared (NIR)**,
- WISE (Cutri et al., 2013), AKARI IRC (Ishihara et al., 2010) and Spitzer (Spitzer, 2015) in **mid-infrared (MIR)**,

- IRAS (Helou & Walker, 1985, IRAS, 1988, Moshir et al., 1989), AKARI FIS (Yamamura et al., 2010) and Herschel (Oliver et al., 2012) in **far infrared (FIR)**,
- ALMA (ALMA, 2015), Planck (Planck, 2011) and WMAP (Gold et al., 2011) in **sub-mm/mm**,
- GB6 (Gregory et al., 1996), NVSS (Condon et al., 1998), FIRST (Helfand et al., 2015), SUMSS (Mauch et al., 2012), WENSS (de Bruyn et al., 1998), 7C (Hales et al., 2007), VLA LFSS (Lane et al., 2014) and a few others in **radio**,
- as well as most important surveys giving **optical images** (DSS I / DSS II and SDSS),
- **proper motions** (Tycho (Høg et al., 2000), USNO, Gaia),
- **variability** (GCVS, NSVS, ASAS, Catalina, LINEAR, Pan-STARRS) and
- **spectroscopic data** (FBS (Markarian et al., 1989), SBS (Stepanian, 2005), Case, HQS (Hagen et al., 1999), HES (Wisotzki et al., 2000), SDSS, 2dF/6dF, CALIFA, GAMA, etc.).

Among the deep fields, HDF N/S, HUDF, CDF N/S, GOODS N/S, and COSMOS are most important. From the years of references, it is obvious that astronomical large-area surveys were carried out especially during the last years and significantly changed our knowledge in all wavelengths.

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 giving high importance to their maintenance, the International Council of Scientific Unions (ICSU, at present: International Science Council, ISC) has created World Data System (WDS) to unify data coming from different science fields for further possibility of exchange and new science projects.

In this paper we give an overall understanding of the astronomical data by coverage along the whole wavelength range and comparisons between various surveys: galaxy redshift surveys, QSO/AGN, radio, Galactic structure and Dark Energy surveys. We describe surveys providing MW photometric data from γ-ray to radio, as well as proper motion, variability and spectroscopic surveys, including objective prism low-dispersion surveys and digital ones.

2. Astronomical Surveys: their importance and main characteristics

The Universe is very big; there are 100s billions of galaxies, each containing 100s billions of stars, nebulae, and other objects. Most of objects are very like each other, and standard approach may be applied to study their average physical properties, structure, and evolution. Classical explanations for stellar configurations, inner structure, stellar atmospheres, and radiation mechanisms have been developed. However, unique objects are needed to study and understand new physical mechanisms, origin and evolution of stars, galaxies, and the Universe as a whole. Some stars have extreme colours, peculiar chemical abundance, emission lines, extended envelopes, some stars are non-optical sources, variables (especially interesting are non-stable ones), binaries (especially interesting are physically connected close binaries), some stars are located in groups and clusters. Some galaxies are peculiar (blue, emission-line, etc.), there are starbursts (SB), active galactic nuclei (AGN), pairs and multiples (especially interesting are interacting ones), mergers, some galaxies have jets, some are non-optical sources, etc. All these peculiar objects comprise typically 5-10% of all observed objects.

It is impossible to study all astronomical objects and one of the main tasks of astronomers is to search and find those peculiar objects that may give more understanding on the physics of objects and phenomena. This task is being achieved by astronomical surveys; having observed large areas of the sky, one can select interesting objects by definite criteria. Selection and application of these criteria defines the value of a survey. Surveys are the backbone of astronomy, and the engine of discovery. They are of cultural importance, because they satisfy the desire to map our surroundings, and give us a feeling for where we live (Lawrence, 2007). Surveys are efficient, because once the sky has been imaged and catalogued, astronomers can do many different experiments using the same database. Surveys data are a resource that supports other astronomy research, e. g. when a γ- or X-ray source is found, one can check whether it is also an IR or radio source, without having to carry out new observations. Some surveys are aimed at mapping a large area of sky, either to build up a large sample to get relevant statistics, or because a large area is being studied like the Milky Way spreading all over the sky. The 20^{th} century technology allowed us to look at the Universe at different wavelengths and many new objects have been discovered at first in non-optical ranges and then identified in optical wavelengths; radio galaxies, quasars, pulsars, cosmic background, molecular clouds, the hot intra-cluster medium, ultraluminous starbursts and AGN (ULIRGs), brown dwarfs, hidden X-ray AGN, etc.

Most important characteristics of astronomical surveys are:

- **Observational method.** Surveys may be imaging (like POSS I and II), photometric, spectroscopic (in this case an objective prism, grism or multi-object spectrographs (MOS) are being used), polarimetric, etc. Some surveys use several modes to combine data and achieve better results. However, this requires more technical efforts and typically is not the case.

- **Sky area.** The selection of the sky area defines the task; e. g. for extragalactic surveys, high galactic latitudes are necessary to skip the heavy galactic absorption. For the galactic surveys, vice versa, definite regions of the Milky Way are being covered.

- **Sky coverage;** depending on this, larger area and more objects may be involved. However, for large sky areas deep surveys are not possible. From this point of view, surveys may be all-sky (totalling 41,253 sq. deg.) and large area ones (a few thousand or a few dozens of thousand sq. deg.) and deep fields (typically less than 1 sq. deg. and often only a few sq. arcmin).

- **Wavelength coverage;** even in optical surveys, wavelength range is important to reveal definite types of objects; e. g. most of the energy of high redshift QSOs is in red and IR part of the spectrum and having observations only in the blue part, one loses many QSOs due to their faintness in that range. Moreover, MW surveys are aimed at discovery of sources in all wavelength domains. In recent SDSS data releases (DR), wavelength range is 3000-10800Å, larger than in all previous optical surveys. Depending on covered wavelengths, the sky may be quite different, therefore this is a rather important parameter.

- **Time coverage.** Typically, most of the surveys make single observations in each field. However, large time domains are necessary for variability studies and repeated observations are being carried out for such purposes. Large time coverage is provided by archival observations that may be used due to digitization of old astronomical plates (variability data are provided by Samus' et al. (2011); Woźniak et al. (2004); Pojmanski (1998); Drake et al. (2014a,b)).

- **Spatial resolution.** The positional accuracy of a survey is derived from its spatial resolution. Recent optical surveys (DSS based catalogues, SDSS, Tycho, etc.) have reached 1 arcsec and better resolution, however in other wavelength ranges there still are technology-based limitations on the accuracy. This also creates inconvenience in cross-correlation of various sources.

- **Spectral resolution.** For spectroscopic surveys, this is one of the most important parameters, as most of the information comes from spectra and more accurate the spectroscopy, more information may be derived. However, high spectral resolutions take longer exposure times, therefore in large surveys, typically low and medium resolutions are being used.

- **Sensitivity.** In optical range, the limiting magnitude of a survey is important to reach fainter objects. Similarly, in other wavelength ranges, the sensitivity (typically given in magnitudes in optical range and UV, mJy-s in IR and radio, or eV-s in high energy astrophysics) defines how deep a survey can reach. In deepest surveys, such as HUDF, 30^m is achieved.

- **Photometric accuracy.** Along with the limiting magnitude or sensitivity, the accuracy of the photometric measurements is rather important. This is the case for estimation of the completeness, derivation of luminosities, colour and variability measurements, etc. Optical surveys reach 0.01^m and better photometric accuracy.

- **Homogeneity.** Any survey needs to be homogeneous, otherwise its value is not maintained. Homogeneous samples of objects or non-optical sources give an important material for statistics and further studies.
- **Completeness.** Based on the homogeneity limit, one may derive the completeness of any survey. As homogeneity, so as completeness gives understanding of the value for a survey. Typically, the completeness of the detection is being considered, however the completeness of classification is also crucial for spectroscopic surveys, which is based on more details, hence it is less than the detection limit.

There are many types of astronomical surveys that may be combined by following criteria:

- **Goals** (sky coverage, discovery of definite objects, etc.). E. g. redshift surveys are devoted to mapping the cosmos in three dimensions. Usually galaxies are the targets, but sometimes these are other objects, such as galaxy clusters or quasars.
- **Object types** (QSOs, AGN, galaxies, blue stars, late-type stars, SNe, variables, exoplanets, etc.).
- **Method** (colorimetric, spectroscopic, multi-band, variability, etc.).
- **Sky area** (all-sky, large area or deep surveys).
- **Wavelength range** (optical, γ-ray, X-ray, UV, IR, sub-mm/mm, radio, combined, MW).

The first systematic redshift survey was the CfA Redshift Survey of around 2,200 galaxies, started in 1977 with the initial data collection completed in 1982. This was later extended to the CfA2 redshift survey of 15,577 galaxies (Huchra et al., 1999). Later on, redshift surveys became most important for large scale structure of the Universe and cosmology.

3. Wide-field telescopes and their discoveries

Historically, astronomical surveys have been carried out with wide-field, mostly **Schmidt telescopes**. Here in **Table 1** we give the list of the largest Schmidt telescopes of the world, most of which at present have historical value. The consecutive columns give: telescope name, correcting lens size in cm, mirror size in cm, focal length in cm, focal ratio, field of view in degrees, plate size in cm, scale in arcsec/mm, location, country, altitude in m, and year of installation.

Table 1. Largest historical Schmidt telescopes.

Telescope name	Corr. cm	Mirror cm	Focus cm	Focal ratio	Field deg	Plate cm	Scale "/mm	Location	Country	Alt. m	Year
Alfred-Jensch	134	203	410	1:3.0	3.4×3.4	24×24	50.3	Tautenburg	Germany	331	1960
Samuel Oschin	122	183	307	1:2.5	6.6×6.6	36×36	67.2	Mt. Palomar	USA	1706	1948
UK Schmidt	122	183	307	1:2.5	6.6×6.6	36×36	67.2	Siding-Spring	Australia	1131	1973
Kiso Schmidt	105	150	330	1:3.1	6.0×6.0	36×36	62.5	Kiso	Japan	1130	1974
ESO Schmidt	102	162	306	1:3.0	5.5×5.5	29×29	67.4	Cerro La Silla	Chile	2400	1969
Jurgen Stock	102	152	301	1:3.0	5.5×5.5	29×29	68.5	Llano del Hato	Venezuela	3600	1976
Kvistaberg Schm.	102	135	300	1:3.0	4.6×4.6	24×24	68.8	Kvistaberg	Sweden	33	1964
BAO 1m Schmidt	102	132	213	1:2.1	4.1×4.1	16×16	96.8	Byurakan	Armenia	1397	1960
Uccle Schmidt	84	117	210	1:2.5			98.2	Uccle	Belgium	105	1958
Hamburg Schm.	81	122	240	1:3.0	5.5×5.5	25×25	86.2	Calar Alto	Spain	2160	1955
Baker-Schmidt	81	91	300	1:3.7			68.8	Bloemfontein	S. Africa	1387	1950
Baldone Schmidt	80	120	240	1:3.0	4.8×4.8	24×24	85.9	Baldone	Latvia	75	1967

Among these, especially **Palomar**, **Siding Spring** and **ESO** Schmidt telescopes are very well known for accomplishment of two Palomar Observatory Sky Surveys (POSS I and II; the two latter telescopes were used for the extension of POSS to the Southern sky). POSS I was carried out in 1948-1958 with Palomar Oschin 1.2m Schmidt telescope in blue and red colours, Kodak 103a-O and 103a-E, respectively. The limiting magnitudes are 21.0^m and 20.0^m. 937 different fields each 6.6°x6.6° were taken and the entire sky above δ −33° was covered. Later on, the southern limit was extended to about −45° (100 more plates); thus the survey as a whole includes 1037 fields. UKST SERC J Southern Survey was carried out in 1975-1987 with UKST 1.2m Schmidt telescope and complemented POSS I. POSS II was accomplished in 1987-2000 for the whole sky in blue IIIaJ, red IIIaF and IR IV-N bands, resulting in limiting magnitudes 22.5^m, 20.8^m and 18.5^m respectively. Both POSS I and II were digitized and Digitized Sky Surveys (DSS I and II) were created, respectively (Lasker et al., 1996, McGlynn et al., 1994). Large informative catalogues were created based on these data (USNO-A2.0, APM, MAPS, USNO-B1.0, GSC 2.3.2)

Byurakan and Hamburg Schmidt telescopes are well known for their spectroscopic surveys; famous Byurakan and Hamburg surveys, respectively (Hagen et al., 1999, Markarian et al., 1989), as well as Hamburg-ESO Survey (HES, Wisotzki et al. (2000)) was done with ESO Schmidt.

One of Byurakan 1m Schmidt telescope's advantages is the presence of its three objective prisms (1.5°, 3°, and 4°), which made possible wide-field spectroscopic observations with various dispersion: 1800 Å/mm, 900 Å/mm, 285 Å/mm near Hγ, respectively. The objective prisms can rotate in the position angle that allows obtaining spectra of any orientation. Markarian survey (or the First Byurakan Survey, FBS) carried out with BAO 1m Schmidt telescope, was one of the most efficient and most important survey in astronomy. It was the first systematic objective-prism survey, the largest objective-prism survey of the Northern sky (17,000 sq. deg) and it was a new method of search for AGNs. It resulted in discovery of 1515 UV-excess (UVX) galaxies, including more than 200 AGN and more than 100 SB galaxies. Markarian survey led to the classification of Seyferts into Sy1 and Sy2 (Weedman & Khachikyan, 1968), the definition of Starburst galaxies (Weedman, 1977), and several other projects, such as FBS Blue Stellar Objects (BSOs, Mickaelian (2008)), late-type stars (Gigoyan et al., 2019), optical identifications of IRAS sources (Byurakan-IRAS Galaxies (Mickaelian & Sargsyan, 2004) and Byurakan-IRAS Stars (Mickaelian & Gigoyan, 2006), BIG and BIS objects, respectively). The Second Byurakan Survey (SBS) was also carried out with BAO 1m Schmidt and was the continuation of FBS to fainter magnitudes (Stepanian, 2005). FBS is now digitized and the Digitized First Byurakan Survey (DFBS, Massaro et al. (2008), Mickaelian et al. (2007)) is available online. It provides 40,000,000 spectra for 20,000,000 objects at high Galactic latitudes. Detailed description of FBS, SBS and DFBS is given in Mickaelian (2014).

As mentioned, HQS and HES also are among most important astronomical surveys. HQS covers 14,000 deg^2 in the Northern sky and HES covers 9,000 deg^2 in the Southern sky. Digitized copies of both HQS and HES are available online. Hundreds of QSOs, other AGN, SB, emission-line galaxies, white dwarfs, cataclysmic variables and other hot stars were discovered using these surveys.

To compare the results obtained by Schmidt telescopes, particularly the spectroscopic surveys, we give in Mickaelian (2016a) a comprehensive table of main characteristics of major low-dispersion surveys and SDSS. Most of them are extragalactic surveys so that mainly high galactic latitudes are covered both in the North and South. This table gives an understanding on various parameters of low-dispersion objective prism surveys and SDSS and proves that many historical surveys are still useful.

In Table 1 we do not give the **Large Sky Area Multi-Object Fibre Spectroscopic Telescope** (**LAMOST**, Xinglong, China at 960 m altitude, installed in 2008), the 4m Schmidt telescope with reflective corrector, with collecting area of 18.86 m^2 and with 20m focal length, as its technology differs from all others and it is a meridian telescope. LAMOST is the first digital Schmidt telescope equipped with 2 4kx4k CCD cameras used in blue and red.

On the other hand, nowadays large (3-4 m) Ritchey-Chretien telescopes give relatively large field of view (up to 1 sq. deg.) and partially substitute Schmidt ones. Some of them are used as survey telescopes. Especially important are ESO's **VST (VLT Survey Telescope)** and **VISTA (Visible and Infrared Survey Telescope for Astronomy)** located in Paranal, Chile. VST is 2.6m optical survey telescope (its primary is 2.61m and secondary is 0.94m); its optical design is modified Ritchey-Chrétien reflector with correctors. VISTA is a 4m NIR survey telescope (its primary is 4.1m and secondary is 1.24m); its optical design is also a modified Ritchey-Chrétien reflector with corrector lenses in camera. Both VST and VISTA also use active optics to map the sky in more details and accuracy.

The future of the historical Schmidt-type and other wide-field telescopes is still disputable. Many of them at present are not used, however some are being modified for further studies. Though many historically important Schmidt telescopes are being closed, anyway such type of optics is extremely useful for new astronomical discoveries. The biggest Schmidt-type telescope, 4m LAMOST is the proof. Moreover, some of the space telescopes have used and now use Schmidt cameras, such as **HIPPARCOS** and **Kepler** telescopes. ESA's HIPPARCOS Space Astrometry Mission was launched in 1989 and operated till 1993 to measure accurate positions and magnitudes and resulted in Hipparcos Catalogue of 118,218 stars and Tycho Catalogue of 2,539,913 stars with the highest accuracy positions and proper motions (Høg et al., 2000). A small 29cm Schmidt camera did all this work. NASA's Kepler mission was launched in 2009 and is aimed at search for habitable planets. Kepler's telescope is a 95cm Schmidt camera with a very wide angle, 105 deg^2. Due to this, it is able to observe 100,000 stars.

Figure 1. World largest historical Schmidt telescopes. Upper row, from left to right: LAMOST, Tautenburg, Palomar and Siding-Spring Schmidt telescopes; bottom row from left to right: Kiso, ESO, Kvistaberg and BAO Schmidt telescopes.

Large Synoptic Survey Telescope (LSST, Vera Rubin Observatory) will be the largest Schmidt type telescope ever built. It is planned for 2021 (fully operational in 2022) and will be installed in Chile. The optical system is three-mirror anastigmat, Paul-Baker / Mersenne-Schmidt design. Its primary mirror will have 8.4 m, secondary, 3.4 m and the tertiary mirror, located in a large hole in the primary, is 5.0 m in diameter. The focal length is 10.31 m. The field of view will be 3.5 deg in diameter, or 9.6 sq. deg. Pixel size will be 0.2 arcsec and the resolution, 0.7 arcsec. Wavelength coverage is similar to SDSS one, 3200-10600 AA. LSST will measure orbits for 100,000 NEOs, it will discover 250,000 SNe per year, its observations will allow building light curves for 2 million QSOs, it will measure proper motions 4 magnitudes deeper than Gaia space mission, and it will construct the dark matter map. LSST will cover 10,000 sq. deg. every 3 nights (the whole sky area in 12 nights!)

and it will be the most powerful optical sky survey for the next two decades.

If comparing historical Schmidt cameras and modern (digital) ones, it is obvious that the beginning of digital era put limitations on fields because of restricted areas of CCDs. Only recently CCDs and their systems began to cover almost similar fields (several dozens of sq. deg.). E. g., POSS fields cover 6.6x6.6 degrees and after digitization and sampling, each pixel is 1.67 arcsec in DSS I and 1 arcsec in DSS II, which means that there are approximately 14kx14k pixels (193 Mpix) and 24kx24k pixels (538 Mpix) in each DSS I and DSS II field, respectively. DFBS sampling is 1.54 arcsec/pix and the field is 4.1x4.1 degrees, so that each DFBS field provides 9.6x9.6 kpix (88 Mpix). Only recently individual CCDs with 15 μm or smaller size pixels are reaching such numbers (the first 4kx4k CCD with 15 μm pixels was produced in 1989), otherwise systems of several CCDs were used to cover enough wide fields comparable to Schmidt telescopes. Anyway, given that CCD has linear response and high quantum efficiency, old observations will not have chance to compete by their quality with modern and future ones. LSST will have the world's largest camera with 3.2 Gpix. One can also compare CCDs of modern best digital photo cameras (40 Mpix) that give detailed images of anything on the Earth but far not enough for astronomical purposes. Here also astronomy proves its modern Big Data nature.

4. Historical Era and Wide-Field Plate Data Base (WFPDB)

Classical Schmidt telescopes gave the vast majority of new astronomical objects making all important discoveries possible. New Schmidt telescopes are now orbiting Earth and Sun and proved huge amounts of data for further astrophysical research. It is pretty obvious that almost all important objects for further astronomical studies have come from wide-field surveys, both colorimetric and spectroscopic. Among the colorimetric surveys, Palomar Observatory Sky Surveys are well known. However, very little information on the nature of these objects may be retrieved from these plates. Spectroscopic surveys give more information about the nature of objects and are much more important, though requiring rather harder work and are thus very rare. Unlike the colorimetric ones, there is no any all-sky spectroscopic survey and only several large area surveys exist.

Before 1609, eye observations and measurements were applied. Then, immediately after the first use of a telescope by Galileo and a number of discoveries, the rapid growth of telescope sizes (both lenses and mirrors) followed. We show in Figure 2 the historical growth of telescopes light collecting area. It is close to logarithmic law.

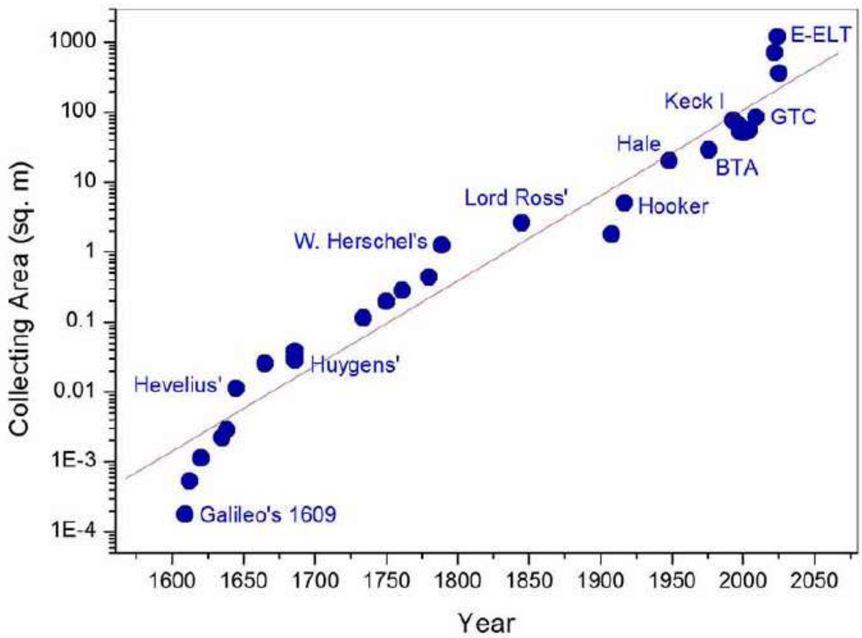

Figure 2. The growth of astronomical telescopes light collecting area since 1609. Largest telescopes of the time and some other important ones are given. The growth is close to logarithmic law.

However, the first telescopes didn't accumulate data, as eye observations were carried out. The telescope era without documentary recording lasted till 1840s, when photography was invented and applied in astronomy. Spectroscopy was an important method introduced in astronomy in early 1800s, even before the photographic era, when Joseph von Fraunhofer used his skills as a glass maker to create very pure prisms, which allowed him to observe 574 dark lines in a seemingly continuous spectrum. Later on, spectrographs were created to obtain and record spectra on photographic emulsion. In early 1900s J. S. Plaskett developed high-quality reflection gratings and more recently, grisms were invented for better quality spectroscopy. **Photographic Era** in astronomy lasted some 150 years and millions of images and spectra were obtained during these years until the beginning of 2000s, when Charge Coupled Devices (CCDs) completely substituted photographic emulsions. The **Digital Era** of astronomy began. However, many astronomers realize and appreciate old archival observations and accumulated data, which are especially useful for variability and proper motion studies (so-called **Time Domain Astronomy**, Mickaelian & Sinamyan (2010), Mickaelian et al. (2011)).

Table 2: Largest ground-based astronomical optical telescopes.

Diam. cm	Name	Year install.	Location	Altit. m	Country
1040	Gran Telescopio Canarias (GTC)	2007	La-Palma, Canarias	2267	Es / US / Mx
982	Keck I	1991	Mauna Kea, Hawaii	4123	USA
982	Keck II	1996	Mauna Kea, Hawaii	4123	USA
920	Hobby-Eberly Telescope (HET)	1997	Mt. Fowlkes, TX	2072	USA
910	South African Large Tel. (SALT)	2003	SAAO	1798	S.Africa / USA
840	Large Binocular Telescope (LBT) 1	2004	Mount Graham, AZ	3170	USA
840	Large Binocular Telescope (LBT) 2	2004	Mount Graham, AZ	3170	USA
830	Subaru	1999	Mauna Kea, Hawaii	4139	Japan
820	VLT Antu	1998	Cerro Paranal, Chile	2635	ESO
820	VLT Kueyen	1998	Cerro Paranal, Chile	2635	ESO
820	VLT Melipal	2002	Cerro Paranal, Chile	2635	ESO
820	VLT Yepun	2001	Cerro Paranal, Chile	2635	ESO
810	Gemini North (Gillett)	2000	Mauna Kea, Hawaii	4214	USA
810	Gemini South	2001	Cerro Pachon, Chile	2715	USA
650	Multiple Mirror Tel. (MMT)	1998	Mount Hopkins, AZ	2616	USA
650	Walter Baade Tel. (Magellan 1)	2002	Las Campanas, Chile	2282	USA
650	Landon Clay Tel. (Magellan 2)	2002	Las Campanas, Chile	2282	USA
605	BTA	1975	Mt. Pastukhovo, Caucasus	2070	Russia
600	Large Zenith Telescope (LZT)	2001	Maple Ridge, BC	395	Canada
508	Hale Telescope	1948	Mount Palomar, CA	1713	USA
425	South. Obs. Astroph. Res. (SOAR)	2002	Cerro Pachon, Chile	2701	Brazil / USA
420	William Herschel Tel. (WHT)	1987	La-Palma, Canarias	2369	UK / Netherl.
420	LAMOST	2008	Xinglong St., BAO	960	China
410	VISTA	2008	Cerro Paranal, Chile	2635	ESO
401	Victor Blanco	1976	Cerro Tololo, Chile	2200	USA

Wide-Field Plate DataBase (WFPDB) was created in Sofia, Bulgaria by Milcho Tsvetkov and colleagues (Tsvetkov et al. (1994); http://www.skyarchive.org) to accommodate all photographic wide-field (> 1°) observations. It contains 414 archives, 2,204,725 plates from 125 observatories obtained with more than 200 telescopes between 1879 and 2002. They include 2,128,330 direct and

64,095 objective prism plates (among them 1874 Markarian Survey and some 600 SBS ones). The biggest archives providing large amount of plates are Harvard – 600,000 plates and Sonneberg – 270,000 plates.

In Table 2 we give the list of the largest ground-based astronomical optical telescopes. One can see how fast the growth of sizes happens. Ex. BAO 2.6m telescope was the 7^{th} largest in the world when it was installed in 1975. However, at present it is not even in the list, being the 45^{th}. However, as most of the big telescopes are in the Western Hemisphere (Hawaii, Chile, etc.), BAO 2.6m is still very important (it is among the 10 biggest ones) for the European-Asian-African region.

5. Multiwavelength era in astronomy and multiwavelength surveys and catalogues

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 (Fig. 3).

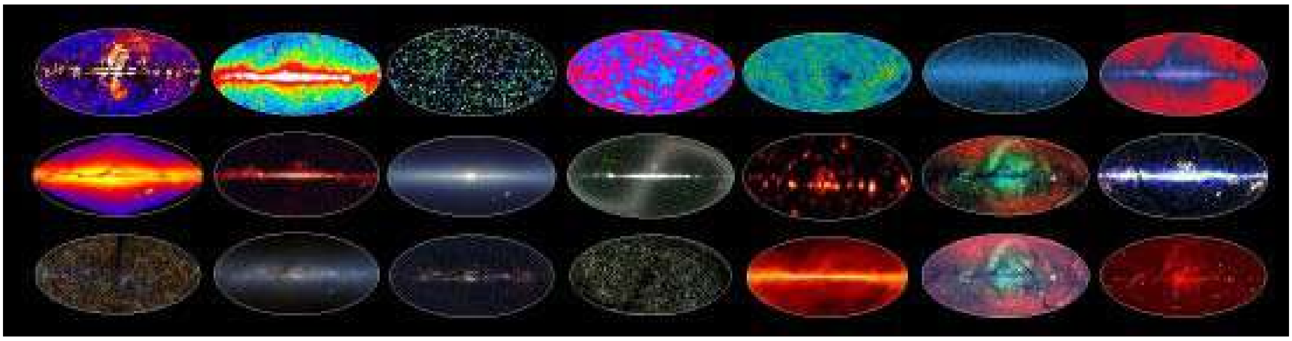

Figure 3. Different views of sky in various wavelength ranges showing the importance of MW studies to have an overall understanding about any given cosmic object and the Universe as a whole.

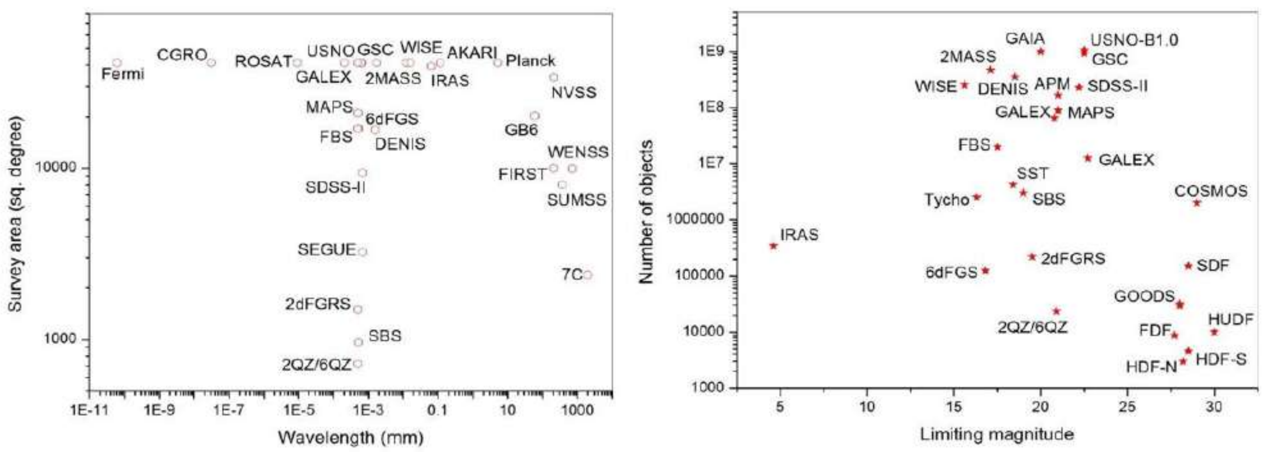

Figure 4. Comparative properties of large surveys. Left: survey area (in square degrees) vs. wavelength (in mm), right: number of survey objects vs. limiting magnitude.

In Mickaelian (2016a,b) 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. To summarize, we give in Table 3 a comparative list of multiwavelength all-sky and large-area surveys.

Table 3: Main data for all-sky and large-area astronomical surveys, as well as some other important projects providing multiwavelength photometric data. Catalogues are given in the order of increasing wavelengths.

Survey, Catalog	**Years**	**Spectral range**	**Sky area** *(deg^2)*	**Number of sources**	**Reference**
Fermi-GLAST	2008-2014	100MeV-300GeV	All-sky	3,033	Acero et al. (2015)
CGRO	1991-1999	20keV-30GeV	All-sky	1,300	Hartman et al. (1999)
INTEGRAL	2002-2014	15keV-10MeV	All-sky	1,126	Bird et al. (2010)
Swift	2004-2008	14-150 keV	All-sky	84,979	D'Elia et al. (2013)
XMM-Newton	1999-2014	0.25-12 keV	Pointed	372,728	XMM-Newton (2013)
Chandra	1999-2014	0.07-10 keV	Pointed	380,000	Evans et al. (2010)
ROSAT BSC	1990-1999	0.07-2.4 keV	All-sky	18,806	Voges et al. (1999)
ROSAT FSC	1990-1999	0.07-2.4 keV	All-sky	105,924	Voges et al. (2000)
GALEX AIS	2003-2012	1344-2831 Å	21,435	65, 266,291	Bianchi et al. (2011)
GALEX MIS	2003-2012	1344-2831 Å	1,579	12,597,912	Bianchi et al. (2011)
APM	2000	opt b, r	20,964	166,466,987	McMahon et al. (2000)
MAPS	2003	opt O, E	20,964	89,234,404	Cabanela et al. (2003)
USNO-A2.0	1998	opt B, R	All-sky	526,280,881	Monet et al. (1998)
USNO-B1.0	2003	opt B, R, I	All-sky	1,045,913,669	Monet et al. (2003)
GSC 2.3.2	2008	opt j, V, F, N	All-sky	945,592,683	Lasker et al. (2008)
FBS	1965-1980	3400-6900 Å	17,056	20,000,000	Markarian et al. (1989)
SBS	1978-1991	3400-6950 Å	965	3,000,000	Stepanian (2005)
HQS	1985-1997	3400-5300 Å	14,000		Hagen et al. (1999)
HES	1990-1996	3400-5300 Å	9,000		Wisotzki et al. (2000)
SDSS DR16	2000-2018	opt u, g, r, i, z	14,555	932,891,133	Ahumada et al. (2020)
SDSS DR16	2000-2018	3000-10800 Å	14,555	4,355,200	Ahumada et al. (2020)
Tycho-2	1989-1993	opt BT, VT	All-sky	2,539,913	Høg et al. (2000)
Gaia EDR3	2013-2020	opt GBP, GRP	All-sky	1,811,709,771	Brown et al. (2020)
DENIS	1996-2001	0.8-2.4 μm	16,700	355,220,325	DENIS consortium (2005)
2MASS PSC	1997-2001	1.1-2.4 μm	All-sky	470,992,970	Cutri et al. (2003)
2MASS ESC	1997-2001	1.1-2.4 μm	All-sky	1,647,599	Skrutskie et al. (2006)
WISE	2009-2013	3-22 μm	All-sky	747,634,026	Cutri et al. (2013)
AKARI IRC	2006-2008	7-26 μm	38,778	870,973	Ishihara et al. (2010)
Spitzer	2003-2009	3-180 μm	Pointed	4,261,028	Spitzer (2015)
IRAS PSC	1983	8-120 μm	39,603	245,889	IRAS (1988)
IRAS FSC	1983	8-120 μm	34,090	173,044	Moshir et al. (1989)
IRAS SSSC	1983	8-120 μm	39,603	16,740	Helou & Walker (1985)
AKARI FIS	2006-2008	50-180 μm	40,428	427,071	Yamamura et al. (2010)
Herschel	2009-2013	55-672 μm	Pointed	340,968	Oliver et al. (2012)
ALMA	2011-2014	0.3-9.6 mm	Pointed		ALMA (2015)

Planck	2009-2011	0.35-10 mm	All-sky	33,566	Planck (2011)
WMAP	2001-2011	3-14 mm	All-sky	471	Gold et al. (2011)
GB6	1986-1987	6 cm	20,320	75,162	Gregory et al. (1996)
NVSS	1998	21 cm	33,827	1,773,484	Condon et al. (1998)
FIRST	1999-2015	21 cm	10,000	946,432	Helfand et al. (2015)
SUMSS	2003-2012	36 cm	8,000	211,050	Mauch et al. (2012)
WENSS	1998	49/92 cm	9,950	229,420	de Bruyn et al. (1998)
7C	2007	198 cm	2,388	43,683	Hales et al. (2007)
VLA LFSS	2007	406 cm		92,965	Lane et al. (2014)

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} Å) 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} Å) 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. Table 4 gives photometric bands of all-sky and large area surveys, as well as some other important projects (e. g. XMM-Newton, Chandra, SST, Herschel, ALMA, Planck, etc.); this way one can get understanding what data at what wavelengths are available and may give new results.

Table 4: Available photometric bands in MW astronomy. All-sky and large-area surveys and some other important projects are given.

Survey, Catalogue	**Photom. band**	**Energy / wavelength**	**Sensitivity**	**Survey, Catalogue**	**Photom. band**	**Energy / wavelength**	**Sensitivity**
Fermi-GLAST	Fermi 5	10-100 GeV		AKARI IRC	L18W	18.0 μm	120 mJy
Fermi-GLAST	Fermi 4	3-10 GeV		WISE	W4	22 μm	6 mJy
Fermi-GLAST	Fermi 3	1-3 GeV		IRAS	25	24 μm	500 mJy
Fermi-GLAST	Fermi 2	0.3-1 GeV		SST	MIPS24	24 μm	0.1 mJy
Fermi-GLAST	Fermi 1	0.1-0.3 GeV		IRAS	60	61 μm	600 mJy
INTEGRAL	IBIS	15keV-10MeV		AKARI FIS	N60	65 μm	3.8 Jy
INTEGRAL	JEM-X	3-35 keV		Herschel	PACS	70 μm	6 mJy
XMM-Newton	Flux5	4.5-12 keV		SST	MIPS70	71 μm	6.0 mJy
Swift	Hard	2.0-10 keV		AKARI FIS	WIDE-S	90 μm	890 mJy
Chandra	h	2.0-7.0 keV		Herschel	PACS	100 μm	6 mJy
XMM-Newton	Flux4	2.0-4.5 keV		IRAS	100	102 μm	1 Jy
Chandra	m	1.2-2.0 keV		AKARI FIS	WIDE-L	140 μm	1.6 Jy
XMM-Newton	Flux3	1.0-2.0 keV		SST	MIPS160	156 μm	80 mJy
Swift	Medium	1.0-2.0 keV		AKARI FIS	N160	160 μm	7.6 Jy
ROSAT	D	0.9-2.0 keV		Herschel	PACS	160 μm	12 mJy
Chandra	s	0.5-1.2 keV		Herschel	SPIRE	250 μm	6 mJy
XMM-Newton	Flux2	0.5-1.0 keV		Herschel	SPIRE	350 μm	83 mJy
ROSAT	C	0.4-0.9 keV		Planck	HFI	350 μm	658 mJy
Swift	Soft	0.3-1.0 keV		SCUBA	450	450 μm	
Chandra	u	0.2-0.5 keV		ALMA	band10	470 μm	3.1 mJy
XMM-Newton	Flux1	0.2-0.5 keV		Herschel	SPIRE	500 μm	103 mJy
ROSAT	A	0.1-0.4 keV		Planck	HFI	550 μm	457 mJy
GALEX AIS	FUV	1539 Å	19.9^m	ALMA	band9	590 μm	1.3 mJy
GALEX AIS	NUV	2316 Å	20.8^m	SCUBA	850	850 μm	
SDSS	u	3551 Å	22.0^m	Planck	HFI	850 μm	289 mJy
POSS I	O	4050 Å	21.0^m	ALMA	band7/8	1.16 mm	140 μJy

Tycho-2	BT	4203 Å	16.6^m	Planck	HFI	1.38 mm	149 mJy
POSS II	j	4680 Å	22.5^m	ALMA	band6	1.70 mm	80 μJy
SDSS	g	4686 Å	22.2^m	Planck	HFI	2.10 mm	169 mJy
Tycho-2	VT	5319 Å	15.2^m	ALMA	band4	2.85 mm	80 μJy
SDSS	r	6165 Å	22.2^m	Planck	HFI	3.00 mm	266 mJy
POSS I	E	6452 Å	20.0^m	WMAP	W	3.20 mm	1 Jy
POSS II	F	6452 Å	20.8^m	ALMA	band3	4.00 mm	50 μJy
SDSS	i	7481 Å	21.3^m	Planck	LFI	4.26 mm	566 mJy
POSS II	N	8060 Å	18.5^m	WMAP	V	4.90 mm	750 mJy
SDSS	z	8931 Å	20.5^m	Planck	LFI	6.81 mm	825 mJy
2MASS	J	1.24 μm	17.1^m	WMAP	Q	7.30 mm	625 mJy
2MASS	H	1.66 μm	16.4^m	WMAP	Ka	9.10 mm	500 mJy
2MASS	K_s	2.16 μm	15.3^m	Planck	LFI	10.56 mm	461 mJy
WISE	W1	3.4 μm	70 μJy	WMAP	K	13.00 mm	500 mJy
SST	IRAC1	3.6 μm	0.6 μJy	GB6		6 cm	18 mJy
SST	IRAC2	4.5 μm	1.2 μJy	NVSS		21 cm	2.5 mJy
WISE	W2	4.6 μm	100 μJy	FIRST		21 cm	1 mJy
SST	IRAC3	5.8 μm	8.0 μJy	SUMSS		36 cm	1 mJy
SST	IRAC4	8.0 μm	9.8 μJy	WENSS		49 cm	18 mJy
AKARI IRC	S9W	9.0 μm	50 mJy	WENSS		92 cm	30 mJy
IRAS	12	11.6 μm	400 mJy	7C		198 cm	40 mJy
WISE	W3	11.6 μm	0.9 mJy	VLA LFSS		406 cm	700 mJy

Thus, 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.

Figure 5 gives the distribution of 58 photometric bands by their effective wavelengths and sensitivity from NIR to radio (most of the 64 bands coming from all-sly or large area surveys are in this ranges): 2MASS, WISE, SST, IRAS, AKARI, Herschel, ALMA, Planck, WMAP, GB6, NVSS, FIRST, SUMSS, WENSS, 7C, and VLA Low-Frequency Sky Survey (LFSS).

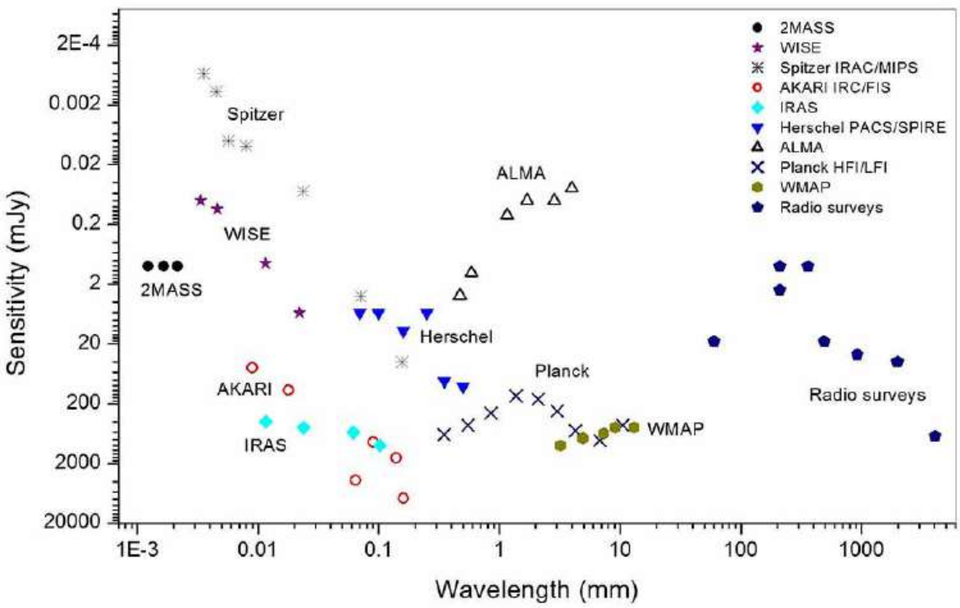

Figure 5. The distribution of 58 photometric bands by their effective wavelengths and sensitivity from NIR to radio. Spitzer and ALMA have especially high sensitivity.

Given that MW data exist in different lists, cross-matching of various astronomical catalogues becomes very important task. Moreover, establishing correspondence between sources revealed in different wavelengths is a tricky work (e. g. Abrahamyan et al. (2015)). Accurate cross-correlations between various MW catalogues are needed to establish genuine counterparts for each object/source. Quick cross-matching is being done for almost all catalogues, however; many objects/sources appear to have false associations, as in crowded regions large contamination with other neighboring objects is happening. Very often individual approach should be applied to such associations. Still, a number of cross-correlation software is in use and is being improved.

6. Big Data Era: numbers 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 biggest catalogues (USNO-B1.0 is the biggest one with 1,045,913,669 objects and GSC 2.3.2 is more accurate with 945,592,683 objects), **SDSS** with its accurate optical images (932,891,133 objects) and spectroscopy providing 10 times more spectra (4,846,156) than available before in astronomy, **WISE** with very accurate positional and NIR/MIR photometric data for 563,921,584 sources that revolutionized astronomy in this wavelength domain. Out of upcoming projects we would like to mention Gaia, LSST and SKA. Table 5 gives the list of the biggest astronomical catalogs.

Table 5: World largest astronomical catalogs.

Survey	**Number of Objects**	**Sky Area**
SuperCOSMOS	1,900,000,000	All-sky
Gaia EDR3	1,811,709,771	All-sky
USNO B1.0	1,045,913,669	All-sky
GSC 2.3.2	945,592,683	All-sky
SDSS DR16	932,891,133	14,555 deg^2
AllWISE	747,634,026	All-sky
2MASS	470,992,970	All-sky

Due to SDSS, the number of QSOs increased up to some 2 million objects (though there is no unique catalog for 2020 (SDSS QSOs from DR16), our estimate is based on the combination of general catalogues of QSOs (Véron-Cetty & Véron, 2010) and QSOs discovered by SDSS DR16) compared to some 30,000 before the SDSS era in 2000, even counting 2QZ/6QZ surveys (Colless et al., 2001, Croom et al., 2004). Due to **Kepler** mission, the number of exoplanets increased to some 4900 (confirmed) and more than 5000 (to be confirmed by spectroscopic observations). More 500,000 QSOs will be provided by Gaia observations, as well as SDSS continues to discover more QSOs in its consecutive surveys (in the next decade, LSST will discover millions of new QSOs). Gaia will also discover some 10,000 or even more exoplanets.

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, 2016a,b). At present astronomers deal with the following numbers in various wavelength ranges (Table 5), and these numbers increase exponentially. 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 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.

As seen from Table 6, 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.

Table 6: Number of catalogued sources at different wavelength ranges giving a comparative understanding about the wavelength coverage of the observed Universe.

Wavelength range	**Major missions, surveys/catalogues**	**Number of catalogued sources**
γ-ray	CGRO, Fermi-GLAST, INTEGRAL, Swift	10,000
X-ray	ROSAT, XMM-Newton, Chandra	1,500,000
UV	GALEX, HST	100,000,000
Optical	SDSS, DSS I, DSS II, HST, Gaia	2,400,000,000
NIR	2MASS, DENIS, HST	600,000,000
MIR	WISE, AKARI-IRC, Spitzer	600,000,000
FIR	IRAS, AKARI-FIS, Spitzer, Herschel	500,000
sub-mm/mm	Planck, WMAP, SCUBA, Herschel, ALMA	200,000
Radio	GB6, NVSS, FIRST, SUMSS, WENSS, 7C	2,000,000

Large astronomical surveys have become one of the most important directions of investigations in our science and they provide the main bulk of information that has been transformed into Big Data and approached astronomy and computer science posing new problems and inquiring new solutions.

As seen, astronomy deals with vast amount of data and big numbers. Table 7 gives some important numbers in astronomy compared to some other numbers known from our everyday life or other sciences. Numbers are given in increasing order from one of the smallest numbers in astronomy (Solar System planets) to the biggest known physical number (atoms in the Universe). Such a comparison allows having an understanding of numbers from different areas and helps remembering them for quick estimations.

Table 7: Numbers from the space +. Most important numbers in astronomy and some other ones for comparison.

Important astronomical and other numbers	**Numbers**
Solar System planets	**8**
Solar System planetary moons	219
Astronomical observatories	500
Trans-Neptunian Objects (TNO) (Dec 2020)	2 500
Discovered Exoplanets (Dec 2020, acc. to Exoplanet.eu)	**4 391**
Discovered Solar System comets	4 894

Astronomical catalogues (in Vizier, CDS, Strasbourg; Dec 2020)	20 398
Number of astronomers in the world (an estimate)	30 000
Number of square degrees in the sky	**41 253**
Human hair (average)	125 000
Astronomical units (AU) in a parsec	**206 265**
Spectral lines in NIST atomic spectra database	227 477
Solar System asteroids (Oct 2020)	998 030
Catalogued X-ray sources	1 500 000
Detected quasars (Milliquas Catalog 6.3, 2019)	1 986 800
Catalogued radio sources	2 000 000
Obtained photographic plates (according to WFPDB)	2 204 725
High and medium resolution spectra in astronomy	7 500 000
ADS abstracts of astronomy/physics papers (Dec 2020)	13 050 798
Seconds in a year	**31 556 926**
Low resolution spectra in astronomy (DFBS, HQS, etc.)	40 000 000
Catalogued astronomical objects	**3 000 000 000**
Age of the Earth in years	4 540 000 000
World population (Dec 2020 estimate)	7 833 166 000
Age of the Universe in years (acc. to Lambda-CDM model)	13 798 000 000
Stars in Our Galaxy	**400 000 000 000**
Galaxies in the Universe	**>500 000 000 000**
Seconds passed after Big Bang	4.35×10^{17}
Total number of animals on the Earth (according to Brian Tomasik)	2×10^{19}
Molecules in 1 cm^3 air (Loschmidt's number)	2.7×10^{19}
Stars in the Universe	$\mathbf{10^{22} - 10^{23}}$
Molecules in the Earth's atmosphere	1.09×10^{44}
Atoms in the Universe	$\mathbf{10^{78} - 10^{82}}$

The information size is given in bytes (B), KB, MB, GB, TB (Terabytes), PB (Petabytes), EB (Exabytes), etc. 1 PB = 1,125,899,906,842,624 or approx. 10^{15} B and 1 EB = 1,152,921,504,606,846,976 or approx. 10^{18} B. 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. Table 8 gives such a comparison. Thus 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.

Table 8: Comparison of information stored in different present and future astronomical surveys or databases and archives.

Surveys, Projects	**Short**	**Range**	**Information Volume**
Digitized First Byurakan Survey	DFBS	opt	400 GB
Digitized Sky Survey (based on POSS)	DSS	opt	3 TB
Two Micron All-Sky Survey	2MASS	NIR	10 TB
Galaxy Evolution Explorer	GALEX	UV	30 TB
Sloan Digital Sky Survey	SDSS	opt	40 TB
SkyMapper Southern Sky Survey	SkyMapper	opt	500 TB
Panoramic Survey Telescope and Rapid Response System	PanSTARRS	opt	∼40 PB
Large Synoptic Survey Telescope, *expected*	LSST	opt	∼200 PB
Square Kilometer Array, *expected*	SKA	radio	∼4.6 EB

The big surveys provided and will provide the following amount of data per year:

- 2008: 20 TB/year (UKIDSS)
- 2010: 100 TB/year (VISTA)
- 2019: 5 PB/year (LSST)
- 2022: 100 PB/year (SKA)

The increase 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.

7. 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.

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. Examples of such software is given in Figure 6 and 7, VOSpec developed by Spanish VO and SED building tool developed by Italian Space Agency (ASI) Science Data Centre, respectively.

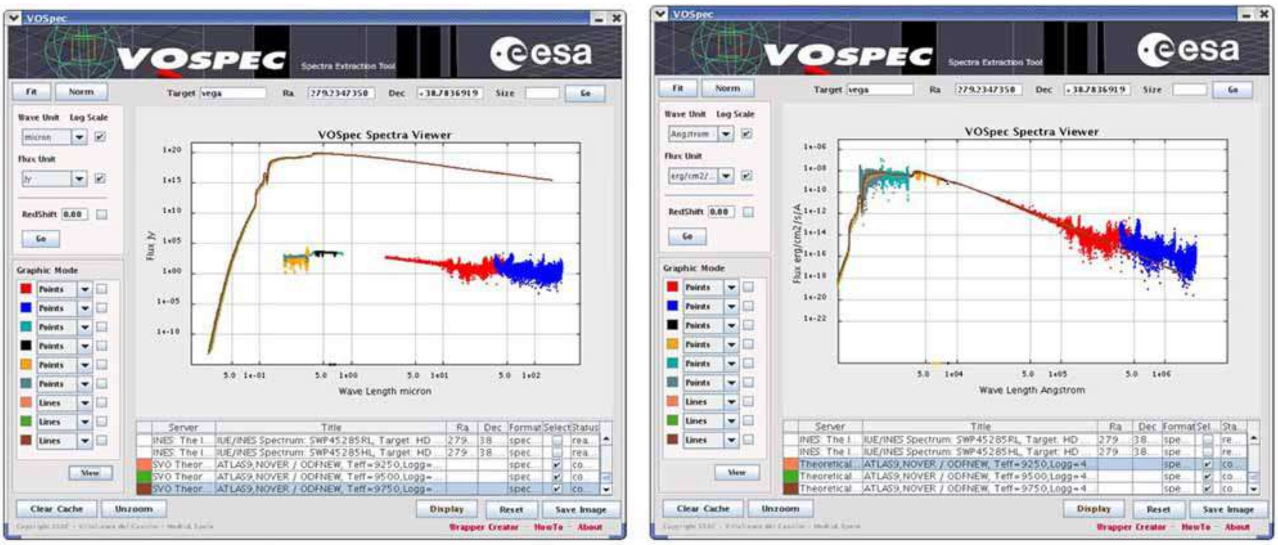

Figure 6. VO software VOSpec allowing superposition and analysis of spectral data coming from different telescopes and different wavelengths, as well as matching with theoretical model curves

Figure 7. SED building software developed at Italian Space Agency (ASI) Science Data Centre. SEDs are given for two Markarian galaxies: Mrk 180 and Mrk 231.

Armenian Virtual Observatory (**ArVO**, https://www.aras.am//Arvo/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. We show in Fig. 8 the logos of DFBS and ArVO.

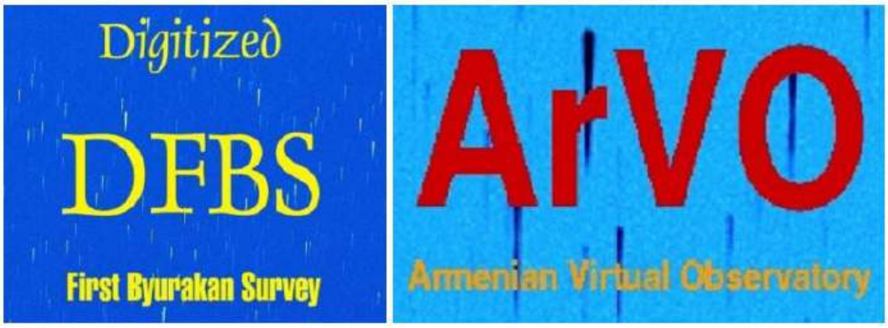

Figure 8. DFBS and ArVO logos.

8. Summary and Conclusions

In 1980s Viktor Ambartsumian was thinking about the growth of astronomical data by comparing the number of published papers. During his young years, 1920s-1930s, he could read ALL astronomical literature. In 1950s, with the growth of these numbers, he could manage to read ALL literature in GIVEN FIELDS, which was especially important for his research. In 1960s-1970s, he was selecting only the MOST IMPORTANT PAPERS from those given fields to read. And in 1980s he could not manage to read even this number of important papers (in 1985, the annual number of published astronomical refereed papers was about 11,000). Ambartsumian concluded that some new approach should be applied and new ways of study of astronomical (and any scientific) literature would be invented. Really, very soon search engines appeared and a new solution was suggested to manage to deal with this large number of information. In Astrophysical Data System (ADS), one can search the whole astronomical (and physical) literature by given keywords (found in the title or abstract), by authors, journals, years, etc. The same situation appeared with astronomical data when working in Internet and later on, by introducing VOs. Astrostatistics is a powerful tool for handling any size of information and providing results on very large datasets.

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, www.eso.org/sci/software/esomidas/) and IRAF (Image Reduction and Analysis Facility, http://iraf.noao.edu/).

In Astrophysics, main results of the 20^{th} 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 has become an important part of astronomical research, without which modern results are impossible.

References

ALMA 2015, http://www.almascience.org/

Abrahamyan H. V., Mickaelian A. M., Knyazyan A. V., 2015, A&C, 10, 99

Acero F., Ackermann M., Ajello M., et al 2015, ApJS, 218, 23

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

Bianchi L., Herald J., Efremova B., Girardi L., Zabot A., Marigo P., Conti A., Shiao B., 2011, Ap&SS, 335, 161

Bird A. J., Bazzano A., Bassani L., et al 2010, ApJS, 186, 1

Brown A. G. A., Vallenari A., Prusti T., de Bruijne J. H. J., Babusiaux C., Biermann M., 2020, arXiv:2012.01533

Cabanela J. E., Humphreys R. M., Aldering G., et al 2003, PASP, 115, 837

Colless M., Dalton G., Maddox S., et al 2001, MNRAS, 328, 1039

Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, MNRAS, 349, 1397

Cutri R. M., Skrutskie M. F., van Dyk S., et al 2003, IPAC/California Institute of Technology

Cutri R. M., Wright E. L., Conrow T., et al 2013, WISE All-Sky Data Release, IPAC/Caltech, VizieR On-line Data Catalog II/328

DENIS consortium 2005, 3rd release, Vizier online data catalogue II/263

D'Elia V., Perri M., Puccetti S., et al 2013, A&A, 551, A142

Drake A. J., Graham M. J., Djorgovski S. G., et al 2014a, ApJS, 213, 9

Drake A. J., Gänsicke B. T., Djorgovski S. G., et al 2014b, MNRAS, 441, 1186

Evans I. N., Primini F. A., Glotfelty K. J., et al 2010, ApJS, 189, 37

Gigoyan K. S., Mickaelian A. M., Kostandyan G. R., 2019, MNRAS, 489, 2030

Gold B., Odegard N., Weiland J. L., et al 2011, ApJS, 192, 15

Gregory P. C., Scott W. K., Douglas K., Condon J. J., 1996, ApJS, 103, 427

Hagen H. J., Engels D., Reimers D., 1999, A&AS, 134, 483

Hales S. E. G., Riley J. M., Waldram E. M., Warner P. J., Baldwin J. E., 2007, MNRAS, 382, 1639

Hartman R. C., Bertsch D. L., Bloom S. D., et al 1999, ApJS, 123, 79

Helfand D. J., White R. L., Becker R. H., 2015, ApJ, 801, 26

Helou G., Walker D. W., 1985, IRAS small scale structure catalog, JPL, Pasadena, NASA

Hey T., Tansley S., Tolle K., (Eds.) 2009, The Fourth Paradigm: Data-Intensive Scientific Discovery

Høg E., Fabricius C., Makarov V. V., et al 2000, A&A, 355, L27

Huchra J. P., Vogeley M. S., Geller M. J., 1999, ApJS, 121, 287

IRAS 1988, Joint IRAS Science Working Group. IRAS PSC, Version 2.0, NASA RP-1190

Ishihara D., Onaka T., Kataza H., et al 2010, A&A, 514, A1

Lane W. M., Cotton W. D., van Velzen S., Clarke T. E., Kassim N. E., Helmboldt J. F., Lazio T. J. W., Cohen A. S., 2014, MNRAS, 440, 327

Lasker B. M., Doggett J., McLean B., Sturch C., Djorgovski S., de Carvalho R. R., Reid I. N., 1996, in Jacoby G. H., Barnes J., eds, Astronomical Society of the Pacific Conference Series Vol. 101, Astronomical Data Analysis Software and Systems V. p. 88

Lasker B. M., Lattanzi M. G., McLean B. J., et al 2008, AJ, 136, 735

Lawrence A., 2007, Astronomy and Geophysics, 48, 3.27

Markarian B. E., Lipovetsky V. A., Stepanian J. A., Erastova L. K., Shapovalova A. I., 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.

Mauch T., Murphy T., Buttery H. J., et al 2012, MNRAS 342, 1117, 2003; Online version 2.1r, 16.02.2012

McGlynn T. A., White N. E., Scollick K., 1994, in 184th AAS Meeting, BAAS 26, 898

McMahon R. G., Irwin M. J., Maddox S. J., 2000, VizieR Online Data Catalog, p. I/267

Mickaelian A. M., 2008, Astron. J. 136, 946. VizieR On-line Data Catalog III/258

Mickaelian A. M., 2014, in Mickaelian A. M., Sanders D. B., eds, Vol. 304, Multiwavelength AGN Surveys and Studies. pp 1–10, doi:10.1017/S1743921314003147

Mickaelian A. M., 2016a, Baltic Astronomy, 25, 75

Mickaelian A. M., 2016b, Astronomy Reports, 60, 857

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

Mickaelian A. M., Sargsyan L. A., 2004, Ap, 47, 213

Mickaelian A. M., Sinamyan P. K., 2010, MNRAS, 407, 681

Mickaelian A. M., Nesci R., Rossi C., et al 2007, A&A, 464, 1177

Mickaelian A. M., Mikayelyan G. A., Sinamyan P. K., 2011, MNRAS, 415, 1061

Monet D., Bird A., Canzian B., et al 1998, USNO Flagstaff Station and Universities Space Research Association (USRA)

Monet D. G., Levine S. E., Canzian B., et al 2003, AJ, 125, 984

Moshir M., Kopan G., Conrow T., et al 1989, IRAS FSC, Version 2.0, NASA

Oliver S. J., Bock J., Altieri B., et al 2012, MNRAS, 424, 1614

Planck 2011, A&A 536, A7, Vizier On-line Data Catalog VIII/88

Pojmanski G., 1998, Acta Astron., 48, 35

Samus' N. N., Durlevich O. V., Zharova A. V., et al 2011, GCVS database, Inst. Astron. and SAI, Moscow

Skrutskie M. F., Cutri R. M., Stiening R., et al 2006, AJ, 131, 1163

Spitzer 2015, http://www.spitzer.caltech.edu/

Stepanian J. A., 2005, RevMexAA, 41, 155

Tsvetkov M. K., Stavrev K. Y., Tsvetkova K. P., Ivanov P. V., Iliev M. S., 1994, in Astronomy from wide-field imaging: Proc. IAU Symp. 161, held in Potsdam, Germany, 1993. Ed. H. T. MacGillivray, E. B. Thomson, B. M. Lasker, I. N. Reid, D. F. Malin, R. M. West, H. Lorenz. Kluwer Acad. Publ., Dordrecht, p. 359

Véron-Cetty M. P., Véron P., 2010, A&A, 518, A10

Voges W., Aschenbach B., Boller T., et al 1999, A&A, 349, 389

Voges W., Aschenbach B., Boller T., et al 2000, MPE, Garching. The ROSAT Faint Source Catalogue

Weedman D. W., 1977, Vistas in Astronomy, 21, 55

Weedman D. W., Khachikyan E. E., 1968, Ap, 4, 243

Wisotzki L., Christlieb N., Bade N., Beckmann V., Köhler T., Vanelle C., Reimers D., 2000, A&A, 358, 77

Woźniak P. R., Vestrand W. T., Akerlof C. W., et al 2004, AJ, 127, 2436

XMM-Newton 2013, XMM-Newton Survey Science Centre Consortium, Leicester, UK, Online catalog 2013yCat.9044....0X

Yamamura I., Makiuti S., Ikeda N., et al 2010, AKARI/FIS All-Sky Survey Point Source Catalogues

de Bruyn G., Miley G., Rengelink R., et al 1998, WENSS Collab. NFRA/ASTRON and Leiden Obs., Vizier online catalogue VIII/62