Modern Trends in Mathematics and Physics
ed. S.S. Tinchev, Heron Press, Sofia, 2009

Astroinformatics: A Synthesis between Astronomical Imaging and Information & Communication Technologies

O. Kounchev[1], **M. Tsvetkov**[2], **D. Dimov**[3], **Ya. Chapanov**[4], **N. Kirov**[1,8], **K. Tsvetkova**[2], **D. Kalaglarski**[5], **S. Christov**[5], **E. Kelevedjiev**[1], **A. Borisova**[2], **J. Goranova**[1,2], **G. Borisov**[2], **R. Bogdanovski**[1], **Al. Kolev**[6], **O. Stanchev**[7], **Al. Marinov**[3], **N. Zlateva**[3], **L. Laskov**[3,8], **G. Marinov**[4]

[1]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 8, 1113 Sofia, Bulgaria

[2]Institute of Astronomy, Bulgarian Academy of Sciences, 72, Tsarigradsko Chaussee Blvd., 1784 Sofia, Bulgaria

[3]Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev St., Block 2 1113 Sofia, Bulgaria

[4]Central Laboratory for Geodesy, Bulgarian Academy of Sciences, Acad. G. Bonchev St., Block 1 1113 Sofia, Bulgaria

[5]National Institute of Meteorology and Hydrology, Bulgarian Academy of Sciences, 66 Tzarigradsko Shaussee Blvd., 1784 Sofia, Bulgaria

[6]"G.S. Rakovski" National Defence Academy, Defence Advanced Research Institute, Military and Staff College, 1504 Sofia, Bulgaria

[7]Department of Astronomy, Faculty of Physics, University of Sofia, 5 James Boucher Blvd., 1684 Sofia, Bulgaria

[8]Department of Informatics, New Bulgarian University, 21 Montevideo St., 1618 Sofia, Bulgaria

Abstract

The newly born area of Astroinformatics has emerged as an interdisciplinary area from Astronomy and modern information and communication technologies, based on the modern Internet developments. Recently, four institutes of the Bulgarian Academy of Sciences launched a joint project called "Astroinformatics" and aimed at the development of the necessary methods and techniques.

As a truly interdisciplinary area Astroinformatics has arisen from the need of ICT methods for preservation and exploitation of the scientific, cultural and historic heritage of astronomical observations. The Wide-Field Plate Data Base is an ICT project of the Institute of Astronomy, which has been launched in 1991, by the Working Group "Wide-Field Sky Surveys" of the International Astronomic Union (IAU) and is unique by its nature at the international level. So far

150,000 plates have been already digitized through several European research programs. As a result, image-data is collected with about 2TB size and it tends to rise up to 1PB. The access, manipulation and science data-mining of such a huge amount of information is a serious challenge for the ICT community and the efforts in this direction are funded by European scientific programs as COST Action 283, FP6 & FP7 of the European Virtual Observatory, the Humboldt Foundation of Germany, and recently, by the Bulgarian National Science Foundation (Ministry of Education and Science of Bulgaria).

1 Introduction

Any area of the scientific knowledge on its way of historical development passes through the quantitative accumulation of facts and data with elements of enthusiasm, spontaneity and indigestion. This development in the pre-Internet era has another negative characteristics – lack of data co-ordination, difficult and hard communication among the scientists working in different countries and institutions. Besides the negative features which we mentioned, this primal development has had its own charm of childish curiosity and naïveté. The next steps of the scientific development in many areas are knowledge systematization and structure shaping. This scheme changed essentially after the appearance of the new methods of information and communication technologies and especially after the discovery of the Internet. The latter created the ground for the development of Virtual Collective Intellect or Cybermind. This gives possibility to the associating units to distribute spontaneously their activities of raw data collection and the parallel data analysis.

First of all, it has become possible now to store the existing data in a digital form. Secondly, due to the new methods for collecting information, the quantity of the new information grows exponentially and its analysis with the old scientific methods and platforms becomes impossible. A possibility has emerged, for the set up of virtual unified platforms, where the physical distances are not important. A new "information topology" has arisen, which represents the intellectual/information closeness for the solution of the scientific problems. One may point out to the increasing interest in the set up and association of virtual platforms/portals as the International Virtual Libraries of books, works of art, data, documented observations, and in particular, the very recent Europeana project (www.europeana.eu) launched in 2006.

Already in the beginning of the application of the more precise and objective photographic methods for astronomical observations of the celestial objects and events, international projects for mapping the sky and measuring the precise stellar coordinates and magnitudes (Carte du Ciel project, Plan of Selected Areas, *etc*.) were established. Such projects were at the grounds of the founding of the International Astronomical Union, coordinating the efforts of all astronomers worldwide. The activities are expanding to all astronomical observations with the build-up of the Center of Stellar Data in Strasbourg (http://cdsweb.u-strasbg.fr), the Harvard Smithsonian Center for Astrophysics and the National Naval Observatory in Washington (www.usno.navy.mil). In the Internet era after the 90s of the last century, new ideas for globalization of

the scientific investigations have appeared, in the frame of GRID-observations and their upgrades as the so called Virtual Portals. One of the major international initiatives was the establishment of The Alliance of International Virtual Observatories (www.ivoa.net), and The European Virtual Observatory (www.euro-vo.org), where the definition "international" has in fact lost its meaning.

2 The Wide-Field Plate Data Base

The participation of Bulgaria in these international initiatives is a matter of high prestige and a necessary element for the participation in astronomical and space investigations. One such program, offered to the Bulgarian astronomical community, is the set up of Wide-Field Plate Data Base worldwide (www.skyarchive.org) [9–12] (Tsvetkov 1991, 1992, 2006, Tsvetkova and Tsvetkov 2006). There are more than 2,200,000 astronomical photographic observations done mainly in professional observatories aimed to investigate the Solar System, as well as our and the other galaxies. These observations have surveyed periodically the sky. With application of the wide-field Schmidt telescopes systematic sky surveys began which contributed to revealing a clearer picture of the Universe. Practically, before the application of the charge-coupled devices (CCD) and orbital multi-wave telescopes the only source of information in Astronomy were the astronomical photographic observations. The main task of the Institute of Astronomy (Bulgarian Academy of Sciences) was collection and analysis of all worldwide wide-field astronomical photographic observations done in more than 200 observatories with more than 400 instruments. The total information recorded on astronomical plates is estimated about 1 PB (=10^6GB=10^9MB=10^{15}Bytes) provided the information is in computer-readable format. It suggests that after the plate inventory and proposed plate catalogue the plates themselves are digitized and online access to the plate image is provided. This task was impossible up to the middle of the 90s of the last century and was only the object of some fantastic emotional models. With the new development of the information and communication technologies and especially with the Internet development including the digitization technologies this problem can be solved completely in the near future. The astronomical photographic observations are a unique source for follow-up investigations of the behavior of the celestial objects for a period of more than 100 years. Thus the build up of such data base would be equivalent to the launching of a virtual telescope working back in time. The determination of the orbits of Solar System bodies which are possible hazardous asteroids (PHA, www.impact.arc.nasa.gov), as well as near Earth asteroids (NEA) are one of the main objectives for establishment of such data base, giving possibility for precise predictions of trajectories and possible impacts (see Clark R. Chapman, David Morrison, Impacts on the Earth by asteroids and comets: assessing the hazard, *Nature* **367** 33-40, 06 January 1994). Such is the case with determination of the orbits of comets coming from the non-predictable Kuiper Belt outside the orbit of Neptune. Only for the last years, by carrying out such observations, the orbits of 4 new small planets were

discovered and determined. This changed the concept of a "planet", in particular the previous understanding that Pluto is a planet was reconsidered. The centers for such observations in Harvard, Rome, Paris, Berlin, operate mainly with a big volume of information and would need an access to these valuable historical observations.

One of the main problems is that the digitized astronomical images have a huge volume (for example: the volume of one digitized plate is from 100 MB up to 1,000 MB, *i.e.* in average 500 MB). The access to such volume of information, as well as its analysis, is now difficult and slow task. On the other hand it is not needed for majority of tasks for analyzing the entire information available on the plate. That is why it is very essentially the digital information to be compressed in such way that the plate information to be stored and retrieved in suitable way after reduction and standardization for usage from many different users.

For these purposes exceptionally useful are the methods for compression, analysis and image proceedings, developed intensively during the last 10-15 years. Their development is due mainly to the Internet progress, modern GRID technologies and GRID-based virtual portals, in particular virtual observatories. Among the compression methods used on a large scale are the recently developed methods based on Harmonic Analysis – such as Wavelets, Frames, Compressive Sensing, *etc*. They found large field of application in one-dimensional signals, as well as in two-dimensional ones. For example, the last versions of JPEG system (JPEG 2000) are strongly based on the theory of wavelets. Recent applications of wavelets to astronomical images have been studied extensively in the works of D. Donoho (Stanford), J.-L. Starck (Paris-Strasbourg) and F. Murtagh (London).*

3 Astroinformatics

The newly born area of Astroinformatics has emerged from a synthesis between Astronomy and modern information and communication technologies, *i.e.* Astroinformatics is a new interdisciplinary area based on advanced Internet applications. Recently, four institutes of the Bulgarian Academy of Sciences (Institute of Mathematics and Informatics, www.math.bas.bg, Institute of Astronomy, www.astro.bas.bg, Institute of Information Technologies, www.iinf.bas.bg, and Central Laboratory for Geodesy, clg.cc.bas.bg) have launched a joint Astroinformatics Project, (www.astroinformatics.eu)

The main problems addressed by the Astroinformatics have the following components:

- Improvement and development of existing algorithms for search in astronomical data bases, as well as in the WFPDB (http://www.skyarchive.org);
- Set up of data base giving interactivity and interoperability in making

*The reader may consult for further information the homepages of D. Donoho (http://www-stat.stanford.edu/∼donoho/), J.-L. Starck (http://jstarck.free.fr/jstarck/Home.html) and F. Murtagh (http://www.cs.rhul.ac.uk/home/fionn/) .

Figure 1. The logo of the Astroinformatics Project, www.astroinformatics.eu.

search with web-base procedures and keywords;

- Improvement and development of new methods for compression, representation and visualization of the images extracted from the photographic plates, which methods are problem-orientated to the theory and practice of astronomy in the frames of the project;

- Set up of Web-access to the digitized original log-books, containing original data of the observers with applying a technology for objects localization in scanned binary images;

- Improvement of the access to the modern astrometric catalogues created and upgraded last decades on the base of the Earth observations and from the Hipparchos and Tycho catalogues.

It is our understanding that the term "Astroinformatics" reflects best the meaning of this new area, which is analogous in sounding with the names of already well established areas as Bioinformatics, Geoinformatics (Geographical Information Systems – GIS), Biomedical Informatics, Neuroinformatics, *etc*., see more in Wikipedia (http://en.wikipedia.org/wiki/Informatics). As it is generally accepted, Informatics is the discipline of structuring, storing, accessing, and distributing information in complex systems.

We have to mention that the term Astroinformatics is in current use by some authors, for defining some special problems in the analysis of astronomical photographs, where some methods of image processing are attracted, see *e.g.* the research of Amos Storkey, at the University of Edinburgh, http://homepages.inf.ed.ac.uk/amos/astronomy.html

Other authors as Kirk Borne (George Mason University), http://cds.gmu.edu/people/faculty/kborne0108.htm, http://classweb.gmu.edu/kborne/*, have the understanding of Astroinformatics mainly as applications of Statistical Methods

*The reader may enjoy reading the paper of K. Borne, which is published online in the book "Next Generation of Data Mining", Eds. Hillol Kargupta, Jiawei Han, Philip S. Yu, Rajeev Motwani, Vipin Kumar, CRC Press, 2008.

and other methods of Computer Science (as Data Mining) to large astronomical databases. He considers Astroinformatics as "classification of real-time astronomical events from very large sky survey projects, data mining related to extragalactic and colliding galaxy research problems, mining of large databases for new knowledge nuggets, development of algorithms for distributed mining of distributed data, novel information retrieval algorithms, scientific database development, archival research with large databases, and science education research specifically focused on inquiry-based science using real science data in the classroom." The same methods of Astroinformatics may be applied to the automated wildfire detection and prediction. In the book "Next Generation of Data Mining", CRC Press, on p. 106, K. Borne defines Astroinformatics as "the formalization of data intensive astronomy for research and education."

4 Constituents of the Astroinformatics Project

As was said above, a successful Astroinformatics project has to comprise a well organized database of scanned images which have to be further subject to fast search, analysis, zoom, retrieval, *etc*. We will provide a short description of the respective methods of ICT to be used in the Astroinformatics project.

In the area of *Image processing*, we have to say that the recent methods of processing, representation and compression have been applied by several authors in treating and analyzing astronomical images retrieved from WFPDB. Let us count a few of these methods: wavelets, frames, curvelets, polyspline wavelets, cf. [5, 8]. The main purpose of the application of these methods is to provide a high compression of the scanned images (the original size of which may vary from 200 Megabyte to 1 Gigabyte). Not less important are the remarkable properties of the wavelet representation (as the strongly local character of the representation) which allow for a fast search of special features having astronomical significance.

In the area of *Image retrieval*, let us mention some recent methods as Context Based Image Retrieval (CBIR) (see [1, 2]) and FANTIR (an approach combining original methods and tools from the areas of Harmonic analysis and Signal processing, Image recognition and Databases, cf. [3]) for the Image processing of the Wide field astronomical plates, and to develop further the recognition and data extraction.

An important point of the Astroinformatics Project will be to feature out new applications of the analysis of the photographic astronomical plates. In particular, nowadays the astrometrical observations (based on photographical plates measurements) from the period 1899-1992 are the only centenary series, used for determination of the variations of the local plumb-lines and Earth rotation. They give us a unique possibility of searching causality between these variations and global changes of the environment, climate, solar activity, ice melting and post-glacial rebound, mean sea level rising, earthquakes, *etc*. Further in this respect, let us mention that the historical observations derived from the Astroinformatics Project may be combined with the modern geodetic, space and geophysical measurements from the space satellites as Gaia and Hipparcos.

An important feature of the Astroinformatics project is the implementation

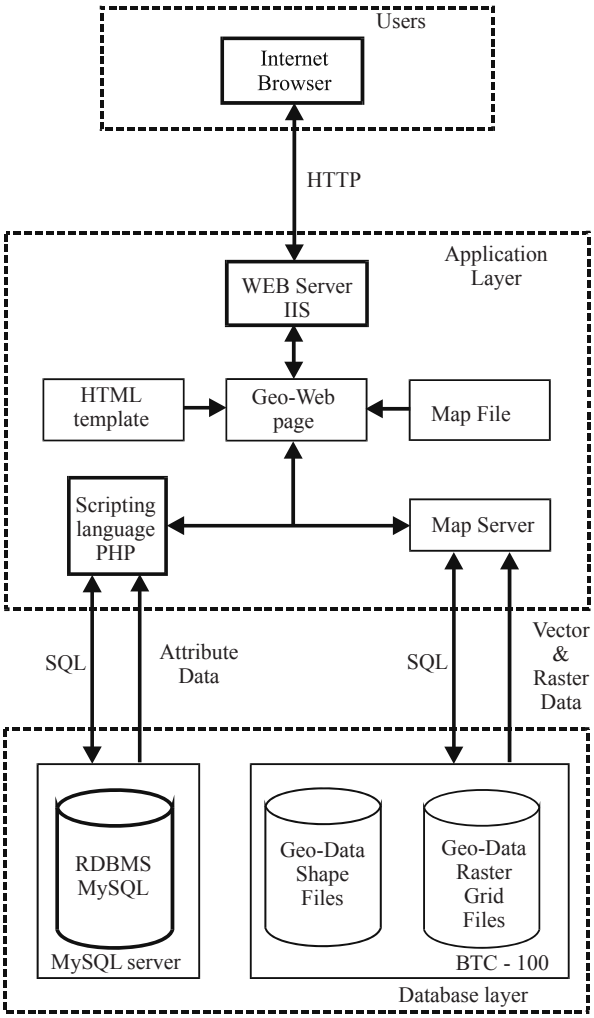

Figure 2. A model structure of an Astro-Web application.

of the database of widefield data plates in a Web-based application. In Figure 2 we present the model software structure of a Web-application representing the implementation of the Astroinformatics Project within LAN or Internet network. It is designed for dynamical interaction by graphics user interface accessing WFPD database. The structure includes a Database Layer, an Application Layer, and Users' Layer.

5 Summary

Let us summarize the main features of the Astroinformatics Project to be carried out at the Bulgarian Academy of Sciences, www.bas.bg.

The main project objective is the usage of ICT methods and instruments for processing, representation and storage of astronomical images in specialized database for efficient data retrieval.

The main tasks to be solved are:

- Preservation, compression and access to wide-field astronomical observations;
- Virtual Observatory ICT standards compatibility, specially for platform independent operability;
- WEB-based search tools for stars in digitized observations;
- Adaptation of methods for image analysis, compression, web-access and data-mining;
- Dissemination among the ICT and astronomical community of the team experience.

The methods which will be used may be characterized as modern ICT for databases management and for WEB-based information system, in particular for signal and image analysis, compression, visualization and recognition, content-based image retrieval, digitization, Fourier analysis, wavelet analysis and time series analysis.

The main expected result is the creation of an ICT system for data preservation and Web-based access to digitized astronomical plates, in view of its future inclusion into the Euro-VO for contemporary astronomical and interdisciplinary research. At the same time different ICT tools will be developed for the research purposes.

Acknowledgments

The project Astroinformatics is supported by Grant of the Bulgarian National Science Foundation under number DO-02-275 and partially by grant number DO-02-273.

Further thanks are due to the Alexander von Humboldt Foundation – Bonn, for the supporting the organization of the Humboldt Kolleg held in Varna, 2008.

About the Authors (The Project Team)

Figure 3. Team of the Astroinformatics Project (2008-2011) with the Bulgarian National Science Foundations. During the seminar at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. Left to right: Aleksander Marinov, Yavor Chapanov, Milcho Tsvetkov, Ognyan Kounchev (project director), Nikolay Kirov, Dimo Dimov, Alexander Kolev, Damyan Kalaglarski, Ana Borisova, Emil Kelevedjiev, Katya Tsvetkova. Picture taken by Nadezhda Zlateva.

References

[1] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain (2000) *IEEE Trans. on PAMI* **22** 1349-1380.

[2] D. Dimov (2007) Rapid and Reliable Content Based Image Retrieval. Proceedings of NATO ASI, Multisensor Data and Information Processing for Rapid and Robust Situation and Threat Assessment, 16-27 May 2005, Albena-Bulgaria, IOS Press, 384-395 (2007) http://www.iit.bas.bg/staff_en/D_Dimov/ASI_Bulg_final_07.pdf

[3] New TO – Software Technology for Fast and Noise Tolerant Image Retrieval based on Graphics Content (FANTIR), Reference OB-0121, May 24, 2004, Technology Offer, IRC-Sofia, 2004; http://www.irc.bg/bg/artShow.php?id=2717

[4] D. Donoho, O. Levi, J.-L. Starck, V.J. Martinez (2002) Multiscale Geometric Analysis of 3-D Galaxy Catalogues, www-stat.stanford.edu/∼donoho/Reports/2002/MGA3D.pdf

[5] D. Kalaglarsky, O. Kounchev (2006) In: *Virtual Observatory: Plate Content Digitization, Archive Mining and Image Sequence Processing*, Eds. M. Tsvetkov, V. Golev, F. Murtagh, R. Molina, Heron Press, Sofia, 245.

[6] O. Kounchev (2001) *Multivariate Polysplines: Application to Numerical and Wavelet Analysis*, Academic Press, San Diego-London.

[7] F. Murtagh, J.-L. Starck (2003) *Astronomical Image and Data Analysis* Springer.

[8] J.-L. Starck and J. Bobin, "Astronomical Data Analysis and Sparsity: from Wavelets to Compressed Sensing", In: Proceedings of the IEEE Special Issue on: Applications of Sparse Representation and Compressive Sensing, http://jstarck.free.fr/IEEE09_sparseastro.pdf

[9] M.K. Tsvetkov (1992) "Wide-Field Plate Database", International Astronomical Union Commission 9 Working Group on Wide-Field Imaging, *Newsletter* No. 2, 51.

[10] M.K. Tsvetkov (1991) "Wide-Field Plate Archives", International Astronomical Union Commission 9 Working Group on Wide-Field Imaging, *Newsletter* No. 1, 17.

[11] M. Tsvetkov (2006) In: *Virtual Observatory: Plate Content Digitization, Archive Mining and Image Sequence Processing*, Eds. M. Tsvetkov, V. Golev, F. Murtagh, R. Molina, Heron Press, Sofia, 10-41.

[12] K.P. Tsvetkova, M.K. Tsvetkov (2006) In: *Virtual Observatory: Plate Content Digitization, Archive Mining and Image Sequence Processing*, Eds. M. Tsvetkov, V. Golev, F. Murtagh, R. Molina, Heron Press, Sofia, 45-53.