Publ. Astron. Obs. Belgrade No. 73 (2002), 149 - 152 *Poster*

THE HYPERLEDA PROJECT EN ROUTE TO THE ASTRONOMICAL VIRTUAL OBSERVATORY

V. GOLEV[1], V. GEORGIEV[2] and PH. PRUGNIEL[3]

[1] *Department of Astronomy and Astronomical Observatory of the St. Kliment Okhridski University of Sofia, P.O.Box 36, BG-1504 Sofia, Bulgaria*
E-mail : valgol@phys.uni-sofia.bg

[2] *Department of Astronomy and Astronomical Observatory of the St. Kliment Okhridski University of Sofia, P.O.Box 36, BG-1504 Sofia, Bulgaria*
E-mail : vladimir@phys.uni-sofia.bg

[3] *CRAL Observatoire Astronomique de Lyon(CNRS: UMR 142), F-69561 St-Genis-Laval Cedex, France*
E-mail : prugniel@obs.univ-lyon1.fr

Abstract. HyperLeda (**Hyper-L**inked **E**xtragalactic **D**atabases and **A**rchives) is aimed to study the evolution of galaxies, their kinematics and stellar populations and the structure of Local Universe. HyperLeda is involved in catalogue and software production, data-mining and massive data processing. The products are serviced to the community through web mirrors. The development of HyperLeda is distributed between different sites and is based on the background experience of the LEDA and Hypercat databases. The HyperLeda project is focused both on the European ***iAstro*** colaboration and as a unique database for studies of the physics of the extragalactic objects.

1. INTRODUCTION

Most of the data from ground-based observatories, and virtually all data from observatories in space, are now carefully collected and preserved for posterity. The volume of data available in the archives is exploding exponentially. Modern ground-based telescopes like the VLT are able to produce around 10 Tb/year whereas the space missions typically produce 10-100 Tb. Besides the quantity, the quality of the data description also improves drastically. Most of these archives are at least partly accessible on-line, but the great diversity of user-interfaces and data management systems limit their use, and attempting to use information from two or more archives together can be exceedingly difficult.

Many people have realized that some form of integrated access is desirable, and that, if powerful data mining facilities were also provided, the scientific gains could be enormous. The name given to this vision is "The Virtual Observatory" (VO). It is worth emphasizing just a couple of positive factors for VO:

⋆ Moore's Law is on our side: computer power, memory size, and disc capacity still seem to be doubling every 18 months.

⋆ Astronomical institutions throughout the world seem extremely willing to collaborate with each other, and to make their services freely available.

Thus a number of collaborations have been formed in support of this aim, and among these is COST 283 or ***iAstro*** (`http://www.iAstro.org`, see Murtagh *et al.*, 2002) which was reported earlier at this meeting (Golev *et al.*, 2002). The HyperLeda project is supposed to play both an important role in ***iAstro*** collaboration and as a unique database for studies of the physics of the extragalactic objects.

2. Objectives of HyperLeda

The HyperLeda project is driven by two astronomical objectives (Prugniel *et al.*, 2002): the study of

- the distribution of galaxies in the Local Universe (distance scale, luminosity function, environment), and
- the evolution of the stellar populations in nearby galaxies.

The principal need for these projects is an access to multi-parametric whole sky samples with extended documentation of the biases.

HyperLeda is based on the background experience of two databases, available through the web: LEDA (see `http://leda.univ-lyon1.fr`), containing homogenized information for nearly three million galaxies, and HYPERCAT (`http://www-obs.univ-lyon1.fr/hypercat`).

The scientific goals are the study of the evolution of galaxies and their environments, their kinematics, stellar population, and nuclear activity, determination of distances and the structure of the Local Universe. The project is focused on the nearby universe, considering it as the reference to investigate the cosmic evolution. Evolutionary studies will be based both on the use of archived data and on the new observations that can be achieved within reasonable amount of telescope time (Prugniel *et al.*, 2001).

HyperLeda develops and maintains a set of tools related to data-mining, massive data processing, data gathering (mainly from published literature and catalogues) homogenization of information and its distribution through web. HyperLeda will allow the external user an access to high quality catalogues, and specialized pipeline to customize the parameterization. Web interfaces from several different sites (mirrors) provide the fastest access to anywhere. Besides the reference data, taken from the astronomical literature and data centers, HyperLeda will make it possible to extract the astronomical information from archives such as DENIS for example.

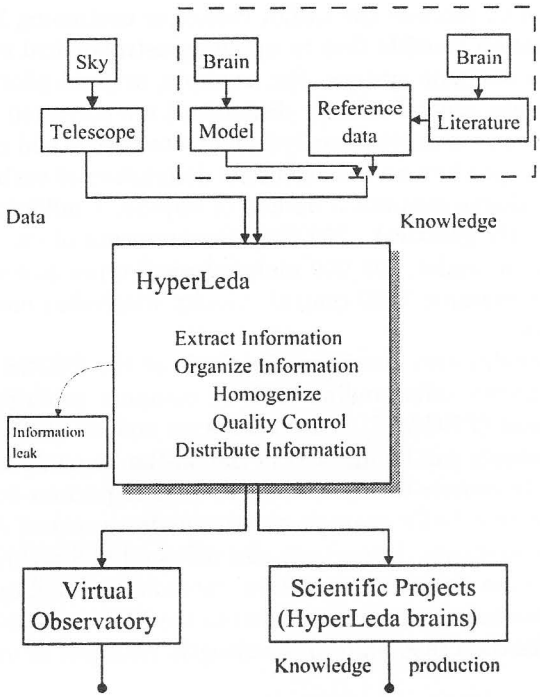

Fig. 1: Information flows in HyperLeda

3. HyperLeda features

HyperLeda implies a collaboration between CRAL Observatoire Astronomique de Lyon, Observatoire de Paris (GEPI), Astronomical Observatory of St. Kliment Okhridski University of Sofia and Sternberg Astronomical Institute at Moscow State University. Solutions proposed by HyperLeda consist in:

(1) Producing 'raw' parameters catalogues compiled from the literature.

(2) Homogenizing these parameters.

(3) Offering on-the-fly functionality to process the data.

(4) Distributing the data through the WEB.

Information flows in the HyperLeda project are shown in Fig. 1. For the different parameters of interest in the frame of our scientific objective we are collecting data from the literature. Each parameter, or group of parameters, is maintained independently by the astronomers contributing to the project. The literature is systematically searched and relevant measurements are converted into an uniform format (eg. by converting the units). These catalogues are automatically asynchronously gathered in the data-flow control center and are checked for consistency. In particular, designation of the galaxies is controlled and cross-identified.

The raw parameters, together with a parameterized description of the measurements in each dataset (series of measurements), are publicly distributed. The raw

catalogues are used to produce the LEDA catalogue containing homogenized parameters. Homogenization consists first in applying astronomical recipes to set all the measurements in a common system. For example, original photometry is converted into Johnson-Cousins, central velocity dispersions are corrected to a standard aperture, and so on. Then, a statistical analysis allows to rescale and combine the different datasets in order to produce an homogeneous description of each galaxy.

The LEDA catalogue contains 3 million of objects, 1 million of them are of confirmed nature (mostly galaxies). 250 000 measurements of cz, 3 million fluxes, diameters and position angles, 225 000 morphological types and several less frequent parameters, as for example 7200 central velocity dispersion measurements or 2000 kinematical profiles.

HyperLeda provides also access to the pixels of the DENIS survey and remote access to other surveys, offers online tools to compute evolutive synthetic spectra of stellar populations (PEGASE) and contains an archive of reference data in FITS format. These resources can be injected in the on-the fly analysis pipeline.

The possibility to process the data on the fly is an important feature of HyperLeda. One of the purposes may be for example to re-normalize a central velocity dispersion in a fiducial or metric aperture. HyperLeda also offers a FITS archive (yet in Hypercat) fed with reference data (images and spectra) intended for calibration, modelling and simulation. The processing pipeline attached to the FITS archive allows to customize the extraction of the data (e.g. shift in wavelength, resample or measure line strength indices).

For example, the pipeline editor command

http://astro.uni-sofia.bg/hypercat/fG.cgi?n=11&z=p&c=i&o=fa:LH_ELODIE/00081

will get as a result the science observation of HD019445 - extracted scan spectrum of exposure 3600 s. Archived data were flat-field corrected , wavelength resampled and flux calibrated above atmosphere.

In the future we will develop the possibility to customize the recipes of data homogenization (ie. give the user the possibility to define his own parameterization) and weighting of individual datasets. The main goal is to make HyperLeda services widely available and easily accessible (possibly) in the framework of some of the ongoing big VO projects.

References

Golev, V., Tsvetkov, M., Murtagh, F., Egret, D. *et al.* : 2002, *Publ. Astron. Obs. Belgrade*, this volume.

Murtagh, F. *et al.*: 2002, *Toward an International Virtual Observatory – Scientific Motivation, Roadmap for Development and Current Status*, ESO/ESA/NASA/NSF Astronomy Conference, June 10 - 14, 2002, Garching, Germany, in print.

Prugniel, Ph. *et al.*: 2002, *Toward an International Virtual Observatory – Scientific Motivation, Roadmap for Development and Current Status*, ESO/ESA/NASA/NSF Astronomy Conference, June 10 - 14, 2002, Garching, Germany, in print.

Prugniel, Ph. *et al.*: 2001, *Mining the Sky, eds. A.J. Banday, S.Zaroubi, & M.Bartelmann*, ESO Astrophysics Symposia, Berlin: Springer – Verlag.