Astronomical Data Analysis Software and Systems XVI
ASP Conference Series, Vol. 376, 2007
R. A. Shaw, F. Hill and D. J. Bell, eds.

Integration of Theoretical Data in the Virtual Observatory

R. Smareglia, P. Manzato

INAF - OATs, Trieste, Italy

C. Gheller

CINECA - Interuniversitary Center for High Performance Computing, Bologna, Italy

U. Becciani

INAF - OACt, Catania, Italy

V. Manna, L. Marseglia, F. Pasian, G. Taffoni

INAF - OATs, Trieste, Italy

Abstract. The Italian Theoretical Virtual Observatory (ITVO) is an ongoing project with the goal of creating a distributed archive of data coming from numerical simulations. It is fully involved in the International Virtual Observatory Alliance (IVOA) standardization process to meet the requirements of theoretical (or simulated) data. The data archive presently consists of the outcome of a set of high end cosmological simulations run with the two popular codes GADGET-2 and Enzo. With respect to the classical way of managing observational data, the ITVO archive introduces the new concept of an evolutional structure. The architecture of the archive is characterized by a core in which the metadata are managed. Then, various interconnected layers can be identified to insert both data and metadata from post-processing procedures which extract information from raw data. These results can be either pre-calculated or generated on the fly at user request. This can be done by exploiting computational systems improving the quality of the information that the ITVO returns to the user. One of the goals of this work is to provide a fully functional interaction with theoretical data within the Virtual Observatory.

1. Introduction

The ITVO database is a multi-level DB where the metadata of the cosmological simulations archive are stored. We focus our work on a set of numerical simulations of galaxy clusters identified at various red-shifts generated by the GADGET-2 code and halo simulations generated by the Enzo code. In particular, for clusters at red-shift zero, we offer the possibility of visualizing the temperature and density profiles and down-loading the maps via a new web interface which is connected to the ITVO database. That interface allows the user to easily search the simulated data by selecting parameters that permit a simple comparison with observations. Another step will be to implement within the database the usage of SNAP (Simple Numerical Access Protocol), which produces on-the-fly the dataset request via the web, for example making a cut-out of the initial simulated box. This latter step will be done using GRID technologies.

2. Storage Simulation

As of today many types of simulations exist: N-body, Mesh, AMR (Adaptive Mesh Refinement), N-body+SPH, N-body+Mesh, N-body+AMR, etc. The scope of these methods is to simulate many different kinds of objects, i.e. dark matter haloes, globular clusters and so on. In this project we deal with two different simulations. The first one is a large cosmological hydrodynamic simulation (Borgani et al. 2004), which used a massively parallel tree N-body/SPH code GADGET-2[1] to simulate a concordance of cold dark matter cosmological model within a box of 192 Mpc. The second simulation is an Adaptive Mesh Refinement (AMR), grid-based hybrid code (N-Body + hydrodynamic) which is designed to simulate cosmological structure formation, using the Enzo[2] code. We stored two CDM simulations made with this code.

3. The ITVO Database and Archive

The variety of the data that we want to store led us to use a relational database, which allows complex queries in a standard language, SQL, also used by many tools developed under the IVOA standards. The database has been designed to store all modern types of cosmological simulations. The database is multi-level: as can be seen in Figure 1, each level holds the data of one step of the whole process. The first level contains the algorithm description, computational and cosmological parameters used for the run, species of matter inserted in the simulation, physical quantities linked to the particles or grid points, and the resolution, redshift and format (HDF5 in our case) of the output file. The next level holds the data of the code used for post-processing, for example the extraction of the clusters and groups of galaxies from the initial box and all the metadata of these new output files, such as virial radius, virial mass, temperature, etc. Another level refers to more refined post-processing: it stores all the metadata of the FITS files concerning the bidimensional maps.

The simulations archive contains the raw data, i.e. the entire simulation outcome, the particle positions and velocities within the simulation box or the physical quantities at each grid point, known as a snapshot at a specific instant. The second level of the archive contains boxes where the code has identified objects of interest; up to now we stored the HDF5 files of the boxes that include the clusters or groups of galaxies. Now the FITS files of all the maps of the clusters at red-shift zero and the relative JPEG files of the profiles have also been saved, but these could be created on-the-fly and removed later when the SNAP protocol allows this.

---

[1] http://www.mpa-garching.mpg.de/galform/gadget/index.shtml

[2] http://www.cosmos.ucsd.edu/enzo/

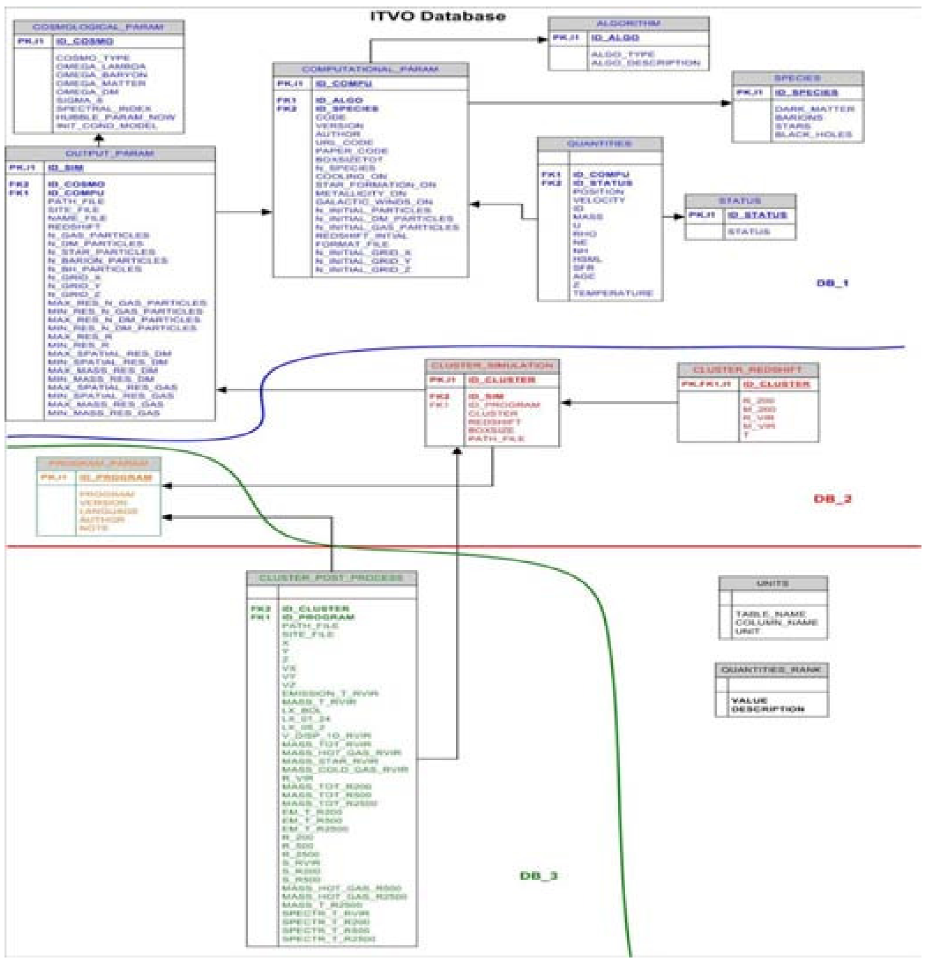

Figure 1. ITVO MultiLayer DB Structure.

3.1. The ITVO Web Interface

At the moment there are three versions of the web interface accessible from the IA2 (Italian Astronomical Archives Centre) site[3]: one to find the raw data box of the simulation, and another to find data of the simulated clusters at several red-shift values. The last one can search and download maps of clusters of galaxies, visualize and print the graphics, visualize the preview of the maps and the header of the FITS file and save as a VOTable the catalog resulting from the database query. In the web interface, it is also possible to customize the SQL query by making a more complex one and/or filtering on different fields, for example the ratio of two selected quantities.

4. SNAP Protocol

SNAP (Gheller et al. 2006) (Simple Numerical Access Protocol, see Figure 2) is a mechanism being currently defined within IVOA to allow access to numerical astyrophysical simulations just as if they were observational data. SNAP is expected to deal with simulations of objects evolving in 3D space. This will allow a spatial subset to be predefined, including various geometries; any simulation that can be expressed as a N-dimensional box could/should also be included. The multilevel database permits access via SNAP to the simulation data at every

[3]http://wwwas.oats.inaf.it/IA2/ITVO/

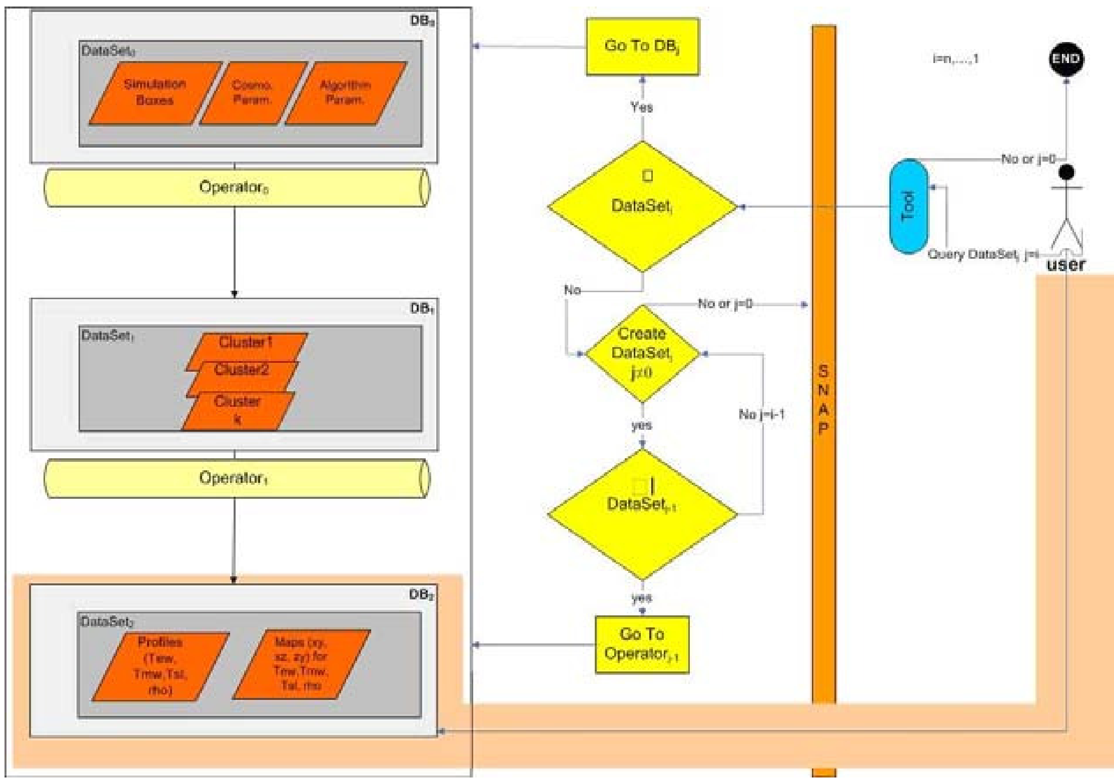

Figure 2. DB Access through SNAP and on-the-fly dataset request.

level. In the future one could create post-processed data of the specific desired sub-box coming from the raw simulated data or create on-the-fly the cluster map or profile. If the request involves a lot of time we might use Grid computation.

5. Conclusion

The goal of creating a distributed database of simulations accessible to the community and its practical implementation point out the need for two more features: evolutional DBs and asynchronous interaction with them. The theoretical VObs will make use of all relevant VObs standards and tools, and will be fully interoperable with VO-complaint databases of observations. This implies good coordination among all the people involved in this project.

Acknowledgments. We are grateful to S. Borgani and S. Ameglio for having provided us with all the files of their GADGET-2 simulation and the codes for the post-processing of these data, and also for their advice on all the questions of scientific nature. Thanks are also due to F. Gasparo for the installation of the DB machines and Oracle 10g software and networking.

References

Borgani S., et al. 2004, MNRAS, 348, 1078
Gheller C., et al. 2006, `http://www.ivoa.net/twiki/bin/view/IVOA/IVOATheorySNAP`