Mem. S.A.It. Suppl. Vol. 13, 62

Memorie della
Supplementi

Data Archiving: experience in the Theoretical Virtual Observatory

P. Manzato[1], M. Molinaro[1], F. Gasparo[1], F. Pasian[1], A. Pietrinferni[2], S. Ameglio[3], G. Murante[4], S. Borgani[3], and S. Cassisi[2]

[1] INAF - Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy, e-mail: manzato@oats.inaf.it
[2] INAF - Osservatorio Astronomico di Collurania, via M. Maggini 64100 Teramo, Italy
[3] Dipartimento di Astronomia dell'Università, via Tiepolo 11, I-34143 Trieste, Italy
[4] INAF - Osservatorio Astronomico di Torino, St. Osservatorio 20, I-10025 Pino Torinese, Italy

Abstract. For archiving and publishing theoretical data it is important to create a suitable structure to store the metadata of the simulation (that must contain all the code parameters used to perform the run) and also all the physical parameters used and at the end also the metadata of the output files; so the database has to contain all the values to rebuilt the production history of a single file. The database must be a relational one in order to permit scientific query on selected parameters to obtain a restrict result. We present the work made for two kinds of astrophysical simulations: cosmological one that contain Gadget, Enzo and Fly simulations and stellar tracks and isochrones performed with the FRANEC code. These two experiences are made inside the VOTECH and VO-DCA EU funded projects.

Key words. Virtual Observatory: archives – Cosmology: simulation – Stars: evolution model – Stars: tracks – Stars: isochrones

1. Introduction

The aim of this work is to follow the Virtual Observatory idea of simplifying the work to astronomers and begin unifying the world of simulations just as it has been made by IVOA with the observational data, see Hanisch & Quinn 2003. So the idea of the Theoretical Virtual Observatory (TVO) was born to develop standard and tools for simulated data in common with the observational ones. The main purpose of the TVO is to create a distributed database of simulated data accessible from anywhere in an easy and transparent way, and to include some services to allow the user to visualize data, download them and extract information from them.

A simulation archive can include an extremely large amount of data; this implies strong difficulties in analysing and moving these kind of data, especially cosmological simulation data. Therefore it is important to develop a system that analyses the data where they are stored. We have built a software layer that can easily allow us to handle data. The challenge is to develop and supply to the community a set of services for data handling

Send offprint requests to: P. Manzato

by providing a user-friendly access to a huge amount of heterogeneous data and also an optimised way to process and analyse these data in a distributed environment. We can refer to three classes of users interested in a data set resulting from a cosmological simulation: the group that performed that simulation; the whole astrophysical community; the non-scientist community, if we consider that a simulation can be used as an auxiliary tool in several educational activities. The web services and the database access are provided through web portals. Such an infrastructure helps avoiding the direct execution of several web services and provides an effective interface to the database service.

2. TVO structure

The TVO schema project structure contemplates a set of distributedarchives and databases and also web portals for the data access that can use many infrastructure and services.
In detail:

1. Archives contain the output files of the simulation;
2. Databases contain the metadata of the simulation that should include all the parameters to perform the run and not only the physical ones;
3. Data Access: it could be performed via Web Portal or Web services or in future via Grid infrastructure, allowing to create on demand new simulated data.

3. Cosmological simulations - ITVO database and web portal

We built a cosmological relational database that has been designed to store all kinds of cosmological simulations. We used an Oracle 10g relational DB, whose query engines allow the user to make complex queries in a standard language, SQL, which is also the standard query language used by many tools developed under the IVOA standards.

As a test-bed we started to deal with three different simulations. The first one is a large cosmological hydrodynamic simulation, see Borgani et al. 2004, which used the massively parallel tree N-body/SPH code GADGET-2.0, see Springel et al. 2001, to simulate a concordance Lambda CDM cosmological model within a box of $192h^{-1}Mpc$. The cosmological parameters assumed were $\Omega_m = 0.3$, $\Omega_b = 0.04$, $H_0 = 70kms^{-1}Mpc^{-1}$ and $\sigma_8 = 0.8$. This simulation regarded star formation, radiative cooling, metal production and galactic wind. It produced 102 snapshots for a total amount of approximately 1.2 TB of raw data.

The second simulation is an AMR (Norman M.L. & Bryan G.L. 1999), grid-based hybrid code (N-Body + hydrodynamic), designed to simulate the cosmological structure formation; it used the Enzo code. We stored two simulations made with this code, both of them simulate a Lambda CDM universe with the following cosmological parameters: $\Omega_m = 0.27$, $\Omega_b = 0.044$, $H_0 = 71kms^{-1}Mpc^{-1}$ and $\sigma_8 = 0.94$.

The third simulation consisted of a set of N-body results, it obtained using the FLY code, see Becciani et al. 2003; typically the N-body simulations were performed on boxes of sizes between 5 and $120h^{-1}Mpc$, with the number of particles varying between 128^3 and 400^3, and for different CDM cosmological models, ranging from $\Omega_{CDM} = 1$ and $\Omega_\Lambda = 0$ to the most popular "concordance" model ($\Omega_{CDM} = 0.3$ and $\Omega_\Lambda = 0.7$).

The ITVO DB, see Manzato et al. 2007 is a multilevel database, as can be seen in Fig. 1. Every level holds the data of one step of the whole data process. The level 0 of the database includes the description of the algorithm, the computational and cosmological parameters used to make the computational run, the species of matter inserted in the simulation, the physical quantities linked to the particles or grid points and also the format (HDF5 and GADGET at the moment in our case), resolution and redshift of the output file. The next level includes the link to the code used for the post-processing, for example the code used to extract the clusters and group of galaxies from the initial box and all also the metadata of these new output file, like virial radius, virial mass, temperature, etc. The last level refers to

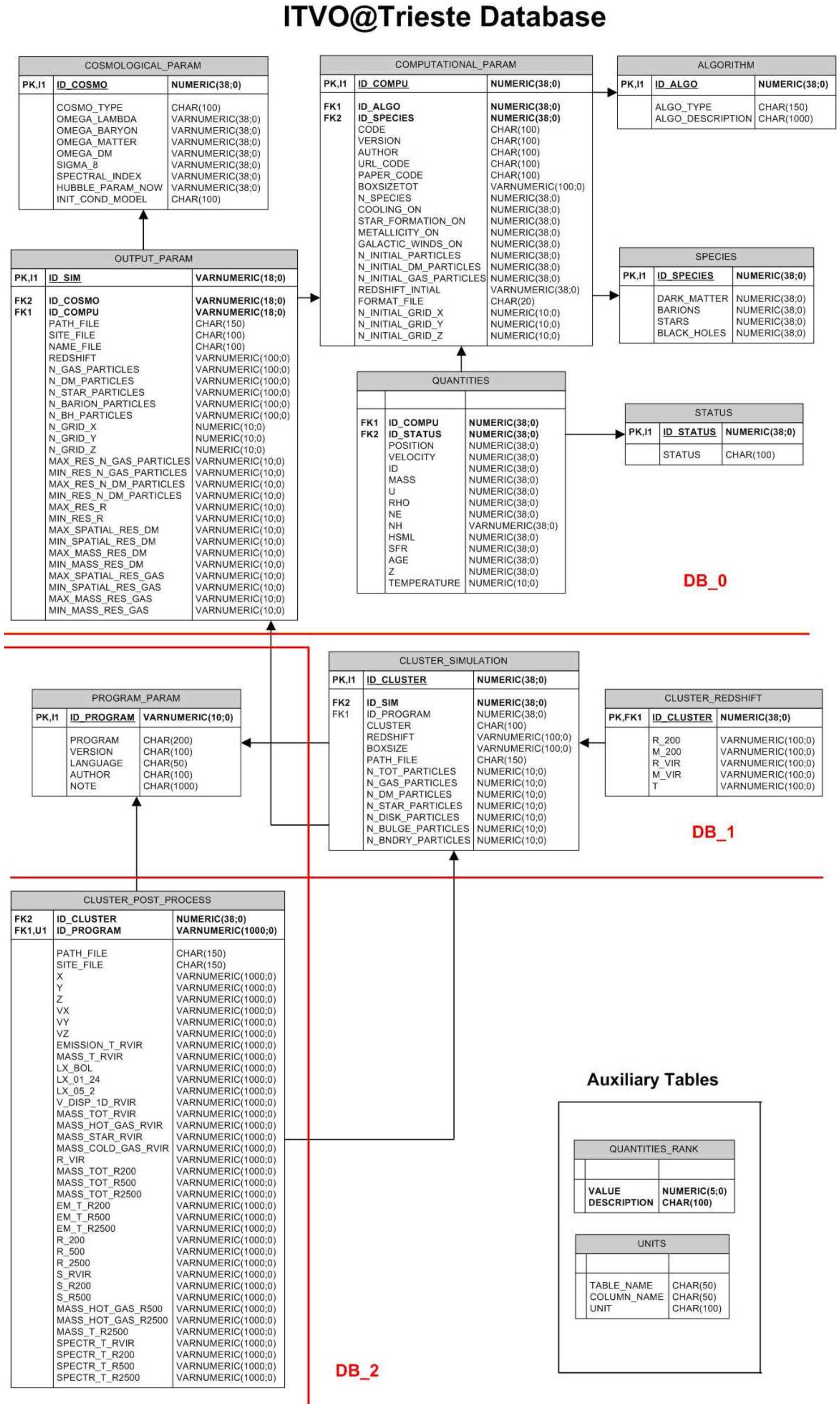

Fig. 1. The ITVO multilevel Database Structure: level 0, level 1 and level 2.

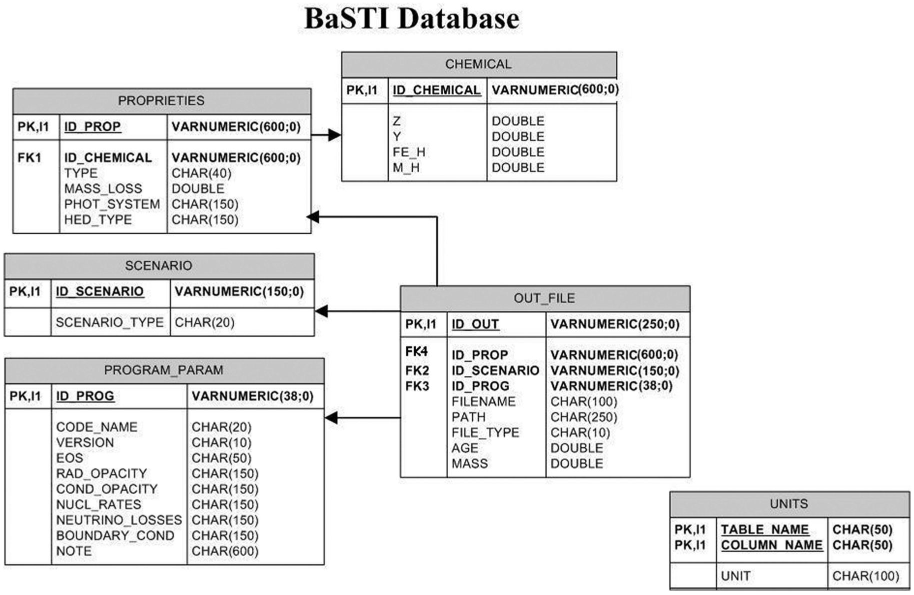

Fig. 2. The BaSTI relational database schema: PK indicates a primary key of a table and FK indicates a Foreign Key to link two tables.

a more refined post-processing: it stores all the metadata of the FITS files concerning the two-dimensional maps.

At present the database access is allowed from three separate levels of web interface accessible from the IA2 (Italian Astronomical Archives Center) web site at the URL, `http://wwwas.oats.inaf.it/IA2/ITVO/`. Furthermore, there are many features described in the Molinaro et al. and Borgani et al. inside these proceedings.

4. Stellar model - BaSTI database and web portal

The database is structured to archive all the parameters regarding a stellar model simulation starting from the initial chemical composition, to the proprieties like: the type of model, the photometric system, the heavy element distribution, the mass loss, and even the type of scenario[1] and all the parameters regarding the numerical evolutionary code linked to the metadata of the simulation output files. This kind of relational database gives the possibility to store and search in an easy manner the data produced by many sets of stellar simulations. It also gives a user-friendly access to a huge amount of homogeneous data like the tracks and isochrones computed by using the FRANEC evolutionary code (see Pietrinferni et al. 2004 and references therein).

The BaSTI web portal is realized in HTML for the static part, while the dynamic part is written in PHP , a server-side scripting language that is especially suited for web development and can be embedded into HTML. The web-server used for our purposes is Apache 2.0 , the most famous open-source HTTP server for modern operating systems. At present the database access is allowed from two different sites:

1. at the IA2 (Italian Astronomical Archives Center) web site at the URL: http://wwwas.oats.inaf.it/IA2/BaSTI/

[1] In this context, with term 'scenario' we refer to the fact that the stellar models are computed under various assumptions about the efficiency of non-canonical physical processes such as core convective overshooting, atomic diffusion, and rotation.

Table 1. The main characteristics of the BaSTI evolutionary models database

mixture	scaled-solar				α-enhanced			
η	0.2		0.4		0.2		0.4	
λ_{OV}	0	0.2	0	0.2	0	0.2	0	0.2
N^O tracks	20	20	40	20	20	20	40	20
$M_{min}(M_\odot)$	0.5	1.1	0.5	1.1	0.5	1.1	0.5	1.1
$M_{max}(M_\odot)$	2.4	2.4	10	10	2.4	2.4	10	10
N^O isoc.	63	44	54	44	63	44	54	44
Age_{min}(Myr)	30	30	30	30	30	30	30	30
Age_{max}(Gyr)	19	9.5	14.5	9.5	19	9.5	14.5	9.5
Photometric system	UBVRIJKL - ACS HST - Ströemgren - Walraven							

2. at the OATeramo at URL: http://albione.oa-teramo.inaf.it/

This functionality permits you to search over a large amount of scientific parameters allowing a simple and direct query to find the data that better satisfy the characteristics of a research typology.

5. Link with the VO

Inside the IVOA we are working to define new standard formats, access protocols and developing tools and web services to make the life of astronomers easier. So we are continuing in the co-operation to develop a DataModel and the access protocol, SNAP, for theoretical data, register the archives and all the services inside the VO registry and increase the number of theoretical data stored inside the TVO. We will also continue developing tools and services able to make an easy comparison between observational and theoretical data.

We are specifically working to trasform the output data in VOTable, that is the VO standard file for tabular data; we are creating and/or modifying tools such as Aladin and VisIVO to enable them to visualize and analyze theoretical and observational data.

6. Conclusions

This work includes the first prototypes to store, access, cutout and analyze the cosmological simulation data in order to reuse the very expensive results of big numerical simulation runs. This work allows the scientists to access and compare theoretical and observational data in an easy and homogeneous way using IVOA standards. Furthermore these databases are continuously update by including additional data and new features to permit analysis and comparison.

Acknowledgements. I am grateful to C. Gheller for his collaboration inside the ITVO projects.

References

Apache Software Foundation, http://www.apache.org/
Enzo Code, http://www.cosmos.ucsd.edu/enzo/
Oracle, http://www.oracle.com
PHP: Hypertext Preprocessor, http://www.php.net/
U. Becciani et al., *FLY. A parallel tree N-body code for cosmological simulations* 2003, CoPhC. Vol. 155, Issue 2, p. 159-179
Borgani S. et al.,2004, MNRAS 348, 1078-1096
Borgani S. et al., 2008 these Mem. S.A.It.
Hanisch R.J. & Quinn P.J., 2003 http://www.ivoa.net/pub/info/
P. Manzato et al., *An archive of cosmological simulations and the ITVO multi-level database* (proceedings EFMCSC series World Scientific, 2007)
Norman M.L. & Bryan G.L., *Cosmological Adaptive Mesh Refinement* (Numerical Astrophysics, eds. S. Miyama & K. Tomisaka, 1999, pp. 19-28)
Molinaro M. et al., 2008 these Mem. S.A.It.
Pietrinferni A., Cassisi S., Salaris M. and Castelli F. 2004, ApJ, 612, 168