Astronomy and Computing 20 (2017) 52–67

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

A web portal for hydrodynamical, cosmological simulations[☆]

A. Ragagnin[b,a,c,*], K. Dolag[c,d], V. Biffi[e,f], M. Cadolle Bel[b,g], N.J. Hammer[a], A. Krukau[a,b], M. Petkova[c,b], D. Steinborn[a]

[a] Leibniz-Rechenzentrum (LRZ), Boltzmannstrasse 1, D-85748 Garching, Germany
[b] Excellence Cluster Universe, Boltzmannstrasse 2r, D-85748 Garching, Germany
[c] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstrasse 1, D-81679 München, Germany
[d] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Strasse 1, D-85748 Garching bei München, Germany
[e] Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy
[f] INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy
[g] Max Planck Computing and Data Facility (MPCDF), Gießenbachstrasse 2, D-85748 Garching, Germany

ARTICLE INFO

Article history:
Received 29 July 2016
Accepted 2 May 2017
Available online 18 May 2017

Keywords:
Cosmology
Galaxy clusters
Online tools
Web application
Cloud computing

ABSTRACT

This article describes a data centre hosting a web portal for accessing and sharing the output of large, cosmological, hydro-dynamical simulations with a broad scientific community. It also allows users to receive related scientific data products by directly processing the raw simulation data on a remote computing cluster.

The data centre has a multi-layer structure: a web portal, a job control layer, a computing cluster and a HPC storage system. The outer layer enables users to choose an object from the simulations. Objects can be selected by visually inspecting 2D maps of the simulation data, by performing highly compounded and elaborated queries or graphically by plotting arbitrary combinations of properties. The user can run analysis tools on a chosen object. These services allow users to run analysis tools on the raw simulation data. The job control layer is responsible for handling and performing the analysis jobs, which are executed on a computing cluster. The innermost layer is formed by a HPC storage system which hosts the large, raw simulation data.

The following services are available for the users: (I) ClusterInspect visualizes properties of member galaxies of a selected galaxy cluster; (II) SimCut returns the raw data of a sub-volume around a selected object from a simulation, containing all the original, hydro-dynamical quantities; (III) Smac creates idealized 2D maps of various, physical quantities and observables of a selected object; (IV) Phox generates virtual X-ray observations with specifications of various current and upcoming instruments.

1. Introduction

Entering the so-called era of "precision cosmology" it becomes more and more clear that a theoretical counterpart in the form of very complex, hydrodynamical cosmological simulations is needed to interpret data from upcoming astronomical surveys and current instruments like PLANCK, South Pole Telescope (SPT), PanStars, Dark Energy Survey (DES), Euclid, LOFAR, eROSITA and many more. Such simulations follow the growth of galaxies and their associated components (like stellar population and central black hole) with their interplay with the large scale environment they are embedded in. Upcoming surveys will map large volumes of the Universe as well as record the birth of the first structures, especially galaxies and even progenitors of massive galaxy clusters at high redshift. In fact, their large potential of determining the nature of dark matter and dark energy comes from being able to map the content and geometry of the Universe over most time in cosmic history. For theoretical models this means that simulations have to cover comparable large volumes, especially to host the rarest, most massive galaxy clusters expected to be the lighthouses of structure formation detectable at high redshift. While the Universe makes its transition from dark matter dominated to dark energy dominated (i.e. accelerated expansion), the objects which form within it make their transition from young, dynamically active and star formation-driven systems to more relaxed and equilibrated systems observed at low redshifts. Those simulations study the internal evolution of clusters of galaxies with respect to the evolution of the cosmological background. They will be essential to interpret the outstanding discoveries expected from upcoming surveys.

☆ Access via https://c2papcosmosim.srv.lrz.de.

* Corresponding author at: Leibniz-Rechenzentrum (LRZ), Boltzmannstrasse 1, D-85748 Garching, Germany.
E-mail address: ragagnin@lrz.de (A. Ragagnin).

http://dx.doi.org/10.1016/j.ascom.2017.05.001
2213-1337/

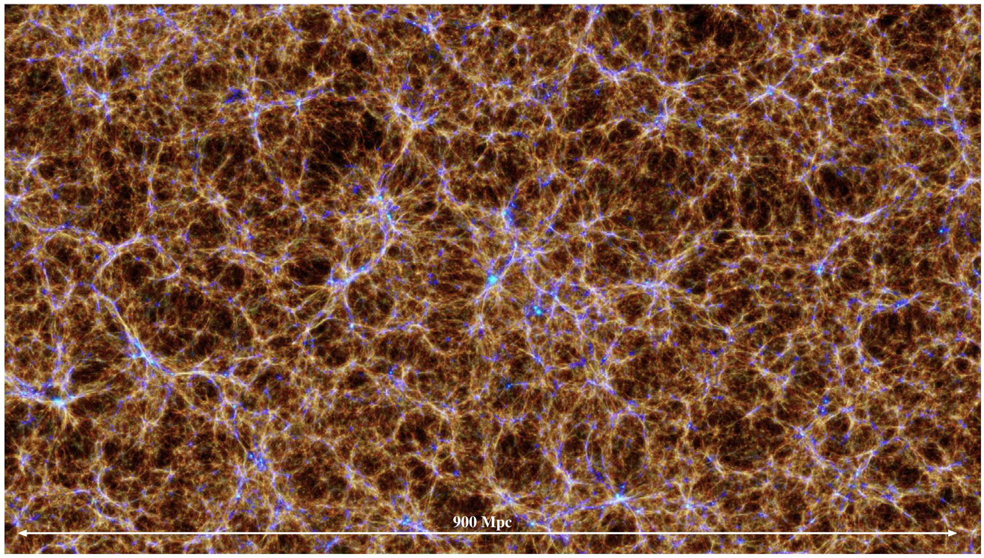

Fig. 1. A visualization of a cosmological large scale structure of the *Box2b/hr* simulation from the *Magneticum* project. This map shows diffuse baryons at $z = 0.2$, colour coded according to their temperature. The visualization is centred on the most massive galaxy cluster in this simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

However, running, storing and analysing such simulations is a challenging task, both from a technical as well as from a collaborative point of view. Recent generations of HPC facilities provided within initiatives like GAUSS[1] or PRACE[2] belong to the first generation of supercomputers which can perform cosmological, hydrodynamical simulations covering both the required large volume and high resolution requirements. Here, the largest simulation performed so far belongs to the *Magneticum* project[3] and follows 2×4536^3 resolution elements over the whole, cosmological evolution of the universe (Bocquet et al., 2016). Such simulations model many more physical processes (star formation, cooling, winds, etc.) than the typical dark matter only counterparts used currently in computational cosmology. These simulations provide a larger set of complex data and can reach several hundreds of terabytes of raw data. Such simulations are performed within large collaborative efforts and results have to be shared with a broader scientific community. A guarantee for a deep scientific impact means that such data are made easily accessible and processable within the individual collaborating groups. It implies that data are stored on the HPC facilities for long periods of time, with the possibility to post-process the full data. In addition, it is important to make such data available to a large astrophysical community and allow the scientists to apply analysis tools via standard interfaces.

In this respect, efforts have been done in the recent years in order to share data sets of various kinds with the community. For instance, the Millennium Simulation Data Archive[4] (Lemson and Virgo Consortium, 2006) is a pioneering work in this field. With the Millennium Simulation Data Archive, the user is able to compose SQL queries over substructure and merger-tree data in order to extract haloes and galaxies from the Millennium Simulation.

Users can also download the raw data files. The Cosmosim.org project[5] allows users to compose additional queries over the list of particles and various post processed quantities (grid cells of density field). The Illustris Galaxies Observatory[6] provides an application programming interface (API) where users can filter galaxies and download particle data from the Illustris simulations. The Australian Theoretical Virtual Observatory[7] (Bernyk et al., 2016) is an online virtual laboratory where users can compose queries and run services on selected objects in the simulation, for instance producing mock observations or extracting light cones.

Section 2 describes data of cosmological simulations and Section 3 describes the currently available infrastructure. In Section 4 we describe how users can interact with the web interface and thereby compose science-driven queries to select objects. Section 5 describes the services currently implemented in the system.

2. The simulations

In this section we present the simulations made accessible by our data centre.

2.1. The magneticum project

The Magneticum simulations[8] (see Biffi et al., 2013; Saro et al., 2014; Hirschmann et al., 2014; Steinborn et al., 2015; Dolag et al., 2015b, a; Teklu et al., 2015; Steinborn et al., 2016; Bocquet et al., 2016; Remus et al., 2016) follow the evolution of up to 2×10^{11} particles in a series of cosmological boxes ranging in size from $(50\ \text{Mpc})^3$ to $(4\ \text{Gpc})^3$. A visualization of the second largest cosmological simulation can be seen in Fig. 1. These simulations were used to interpret Sunyaev–Zel'dovich (SZ) data from PLANCK (Planck Collaboration et al., 2013) and SPT (McDonald et al., 2014) as well as to predict cluster properties in X-rays for future missions such as Athena or Astro-H (Biffi et al., 2013). The first mock observations for the eROSITA cluster working group and the Athena+ white book were also produced based on these simulations. Other scientific goals that were achieved with these simulations included studying the properties of the intra cluster medium (ICM) in galaxy clusters (Dolag et al., 2015b) as well as predicting the multi wavelength properties of the Active Galactic Nuclei (AGN) (Hirschmann et al., 2014; Steinborn et al., 2015). The large dynamical range probed by the combination of resolution

1 https://gauss-allianz.de/
2 http://www.prace-ri.eu/
3 http://www.magneticum.org
4 http://wwwmpa.mpa-garching.mpg.de/Millennium/
5 https://www.cosmosim.org/
6 http://www.illustris-project.org/galaxy_obs/
7 https://tao.asvo.org.au/tao/about/
8 http://www.magneticum.org