

# Simulation post-processing on the Grid

*H.-M. Adorf, AstroGrid-D@MPA*

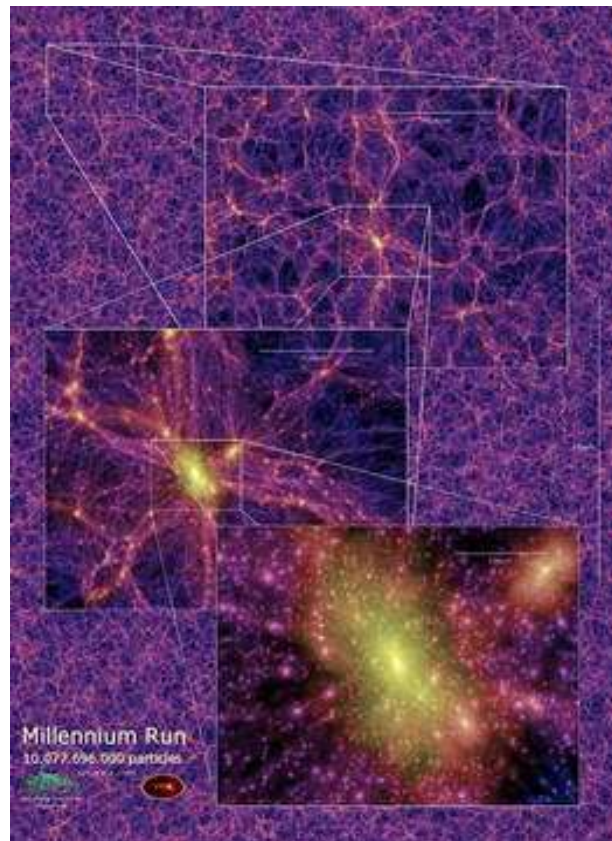
*2005-10-31, V 1.2*

**Abstract:** An AstroGrid-D project is outlined which consists of grid-enabling the on-demand post-processing of the output data from cosmological simulations, in particular the 25 Terabyte Millenium Simulation run of the Virgo consortium. Up to  $10^5$  files with a total size of 0.2 to 1 Terabyte need to be managed in a single post-processing workflow comprising about half a dozen individual stages. The grid-enabled Process Coordinator (ProC) scientific workflow engine shall be used for the design and execution of post-processing workflows. In addition to simulation post-processing, the GADGET-2 simulation code itself shall be invocable on demand.

## Background

The cosmological simulation code GADGET-2 is a new massively parallel TreeSPH code, capable of simulating a collisionless fluid with the N-body method, and simulating an ideal gas by means of smoothed particle hydrodynamics (SPH) (Springel, 2005, Springel, 2005).

By far the largest output data set produced by GADGET-2 is the Millenium Simulation run of the Virgo Consortium (Anonymous, 2005, Springel, White & Lemson, 2004, Springel, White, Jenkins, et al., 2005). These cosmological simulations have consumed about 300,000 CPU hours. The raw data sets shall be published in 2006, whereas mock galaxy catalogues shall already be published by the end of this year (2005). The data sets are large, and post-processing potentially entails high demand on processing power, memory, disk space, and network bandwidth.



In order to allow comparisons between the simulations and observational data, the simulation output needs to be further processed. Since the details of simulation post-processing depend on the science goals, it cannot be carried out in a generic fashion.

Post-processing requires chaining about half a dozen separate tasks, which may produce up to  $10^5$  files. Obviously, managing such a large number of files is tedious and error-prone, and often students, when supervising the post-processing chain, are overwhelmed by the large number of files, particularly when something went wrong.

## **Use cases**

### ***Post-processing of simulation data***

The AstroGrid-D project described here aims at grid-enabling an on-demand post-processing of cosmological simulation data produced by the GADGET-2 simulation code. In view of the complexity of simulation post-processing, we plan to use the power of a grid-enabled Process Coordinator (ProC) for workflow design, job submission and control. The Data Management Component (DMC) software shall be used for the management of the intermediate and final data products.

To this end the user shall be given a predefined template ProC workflow that organizes the data flow through the various post-processing stages. The user may choose parameters and submit the workflow to the workflow execution system. The execution will take place on the underlying Grid system, exploiting parallelism, wherever possible.

The Grid system may be a local cluster managed by a resource management system such as the Portable Batch System (PBS) or the Sun Grid Engine (SGE), an institute grid, or a global grid accessible via the Globus Toolkit 4 or the UNICORE grid middleware.

### ***Running the simulation code***

The simulation code itself is highly parallelizable, and will separately be grid-enabled by the DEISA supercomputing project (Pringle, 2004). In an extension to the originally proposed project, we plan to allow the invocation of a grid-enabled GADGET-2 code for smaller simulations.

In this scenario, the post-processing workflow may be triggered by each simulation snapshot as released by the running GADGET-2 process. This will allow monitoring the simulation process while it is happening, and facilitate diagnosing problems during the run.

It is currently unclear whether the grid-enabled GADGET-2 code maintained by DEISA will be invocable, or whether another way needs to be envisaged.

## **Datasets**

The input to simulation post-processing may be any of the Gadget binary output files, which are structured depending on the situation. These files comprise

- galaxy clusters,
- dark matter position files, and/or
- hydro-simulation files.

## **Location**

The output of the Millenium Simulation run comprises about 25 Terabyte, and is preserved in a robotized long-term mass storage system at the Rechenzentrum Garching (RZG).

## **Storage system**

Access to the data proceeds via an unspecified network protocol. Data can be transferred from tape to local disk with 6 Mbyte/sec. It is possible to stage the data on a RAID-server at RZG or at MPA.

A distributed storage of the simulation data on the D-GRID is conceivable.

## **Storage format**

Currently, the output data sets are file-oriented.

In the future metadata shall be administered in a database using the DMC.

## **Structure**

A post-processing request may comprise 20 to 50 datasets. Each dataset contains typically 10 to 20 Gigabyte. Thus an individual post-processing request involves the processing of 200 Gigabyte to 1 Terabyte of data.

The workflow will process  $10^2$  to  $10^5$  files, typically  $10^3$ . A single file contains 1 Megabyte to 2 Gigabyte with  $10^5$  to  $10^{10}$  rows (records). There are different file types:

- primary data for simulation output, and
- tables of simulation “particles” (dark matter, galaxies, ...) + metadata.

Post-processing produces other output files: e.g.

- an identification of particle groups,
- “clusters of galaxies” with back-references to the simulation particles, or
- properties of galaxies and galaxy clusters.

Post-processing requires to read in the original simulation output at the very beginning. It may be required to read the original output at a later post-processing stage, so depending on the workflow, the input data may be deletable or not.

## **Format**

The output of the simulation is in form of files that are usually formatted in the proprietary Gadget binary format (preferred).

As an alternative, NASA's HDF5 standard binary format may be used. The latter is very flexible, and machine independent, but requires complex I/O. Within this project the compatibility between data structures of HDF5 and the DMC will be investigated.

no exchange between different simulation code outputs [???

## **Read & write access**

The simulation data is usually not modified. It is „write once, read many-times” (WORM) data.

There are updates to the data once a month. [???

Code can read the data via access libraries (C, IDL, also if necessary C++, Fortran).

Status: personal production of simulations [???

Future: exchange of simulation [???

Post-processing occurs a few times per month.

Caching of intermediate products and final products is possible via ProC. Caching can be defined by the user or by the system manager through switches in the pipeline. There is an envisaged option to cache only the metadata.

Parts of the post-processing is standardized and it makes sense to cache the output; other parts of post-processing are special for an individual researcher, and there is no need to cache the intermediate or final data products.

### **Software resources**

The simulation software comprises the following modules:

- GADGET-2 simulation code;
- GridGADGET at AIP (?)

The post-processing software comprises the following modules:

<b>Module</b>	<b>Functionality</b>
Friends of friends (FOF)	for the identification of groups
Subfind	find dark matter substructure
Basetree	first phase of a two-stage process for finding merger trees; constructs layer between subsequent time steps; tests all halos in same time steps
Halotrees	second phase; self-contained data structures for merger trees; separation, constructs complete history; e.g. forest of $2 \times 10^6$ trees
Lgalaxies	applicable to each tree; attaches physical galaxy properties; possibly database query
Splotch	ray tracing for hydro-simulations
Smag	generates physical maps smooth particle hydrodynamics (SPHs)
Set of IDL scripts	Visualization

In addition, tools exist such as viewers (e.g. the public tool CosmoLab, Bologna). One program produces FITS tables and images.

A compilation of the various data formats and types used by these modules is required.

### **Lifecycle for metadata**

Simulation data comes with associated metadata. Data and metadata are synchronously generated. Metadata may be preserved, while intermediate data will often be deleted (in order to save disk space).

### **Availability of metadata**

Metadata is dynamically generated, and is currently be stored by the modules along with the data. Most of the metadata, if not all, is represented by key-value pairs. For instance, the maps are formatted as FITS files with FITS headers storing the metadata as key-value pairs.

The metadata precedes the particle data files. The simulation parameters are described in input files; they are not necessarily stored along with the simulations.

In the future the DMC shall store the metadata in the underlying database.

## **Metadata schema**

The schema needs to be extensible in order to accommodate future changes in the simulation and/or post-processing code. The DMC is capable of handling extensions.

## **Application scenarios**

Post-processing of simulation data sets should be done within the ProC.

The ProC should have restart capabilities, in order to be able to resume processing after a serious error has occurred. This requires coordination of ProC and DMC, and probably user interaction. This function is not yet available.

## **References**

- Anonymous (2005): "World's Largest Supercomputer Simulation Explains Growth of Galaxies", **2005**,
- Pringle G. J. (2004): "DEISA - Distributed European Infrastructure for Supercomputing Applications",
- Springel V. (2005): "The cosmological simulation code GADGET-2", *Mon. Not. Royal Astr. Soc.* (**submitted**),
- Springel V. (2005): "GADGET-2: Galaxies with dark matter and gas interact - A code for cosmological simulations of structure formation",
- Springel V., S. White and G. Lemson (2004): "Analysing and Publishing the Virgo millenium simulation on the D-GRID", 3.
- Springel V., S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg and F. Pearce (2005): "Simulations of the formation, evolution and clustering of galaxies and quasars", *Nature* 629-636.