Towards multidimensional VO archives in the exascale era

Ruiz, J.E. [1], Santander-Vela, J.D. [2,1], Verdes-Montenegro, L. [1]

(1) Instituto de Astrofísica de Andalucía – CSIC
(2) European Southern Observatory

Abstract

Because of the size and complexity of the next generation of observational datasets, data providers must store huge volumes of data, but at the same time should supply on-line processing and analysis services that will provide on-the-fly generated data. At this moment some efforts inside the Virtual Observatory community are being steered towards the achievement of a second generation of protocols that would allow for *in-silico* computational standardized tasks. These services can be reused as components for internet-based workflows that capture and preserve the scientific methodology. Scientists should be able to easily build, customize and share workflows among the community allowing the reproducibility of the results and fostering collaborative work in the interest of scientific discovery.

Virtual Data Products

Among the main issues and challenges that need to be addressed for the next generation of astronomical archives are: a data description model for radio-interferometrical and multidimensional data in the Virtual Observatory (VO), a format unification for science-ready data files, and mature standard VO protocols for discovery and data access and transmission.

Virtual data should be generated at access time providing not only subsets, projections and dimensional reductions of complex observations, but also derived products issued from tailored sophisticated analysis tasks deployed as standard VO services.

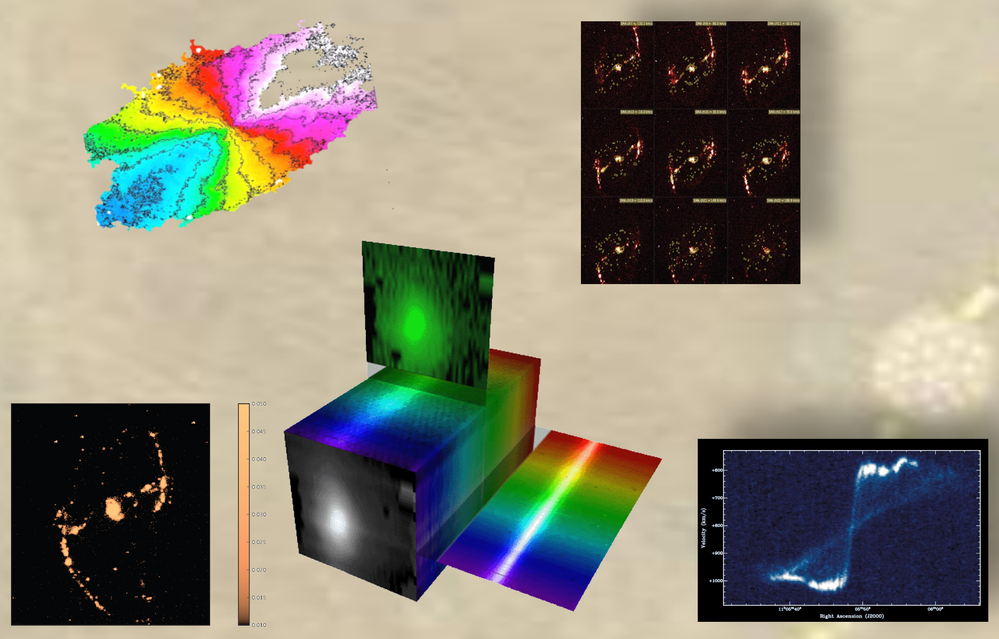

Figure 1. Different virtual data products of NGC1530 generated as cutouts, 2D slices, projections, dimensional reductions and axis transformations over a multidimensional dataset.

Infrastructure

In order to process the expected torrent of raw data in real time, full exploitation of the GRID facilities will be considered in a new innovative way, as those related with global sensor networks and stream computing. This will need the GRID infrastructure to realize its full capabilities, providing not only high performance distributed computing, but also distributed data storage and innovative retrieval technologies when data holdings reach the exascale.

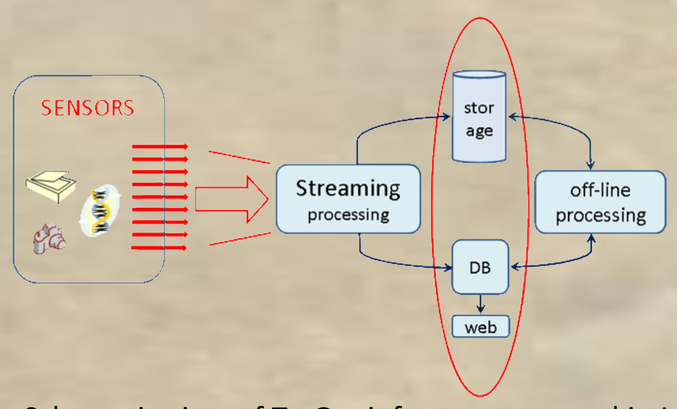

Figure 2. Schematic view of TarGet infrastructure used in LOFAR telescope for sensor network data streaming processing and off-line processing.
http://www.rug.nl/target

Scientific Workflows

The quantitative leap in volume and complexity of the next generation of archives will need analysis and data mining tasks to live closer to the data, in computing and distributed storage environments (GRID), but they should also be modular enough to allow customization from scientists and be easily accessible to foster their dissemination among the community.

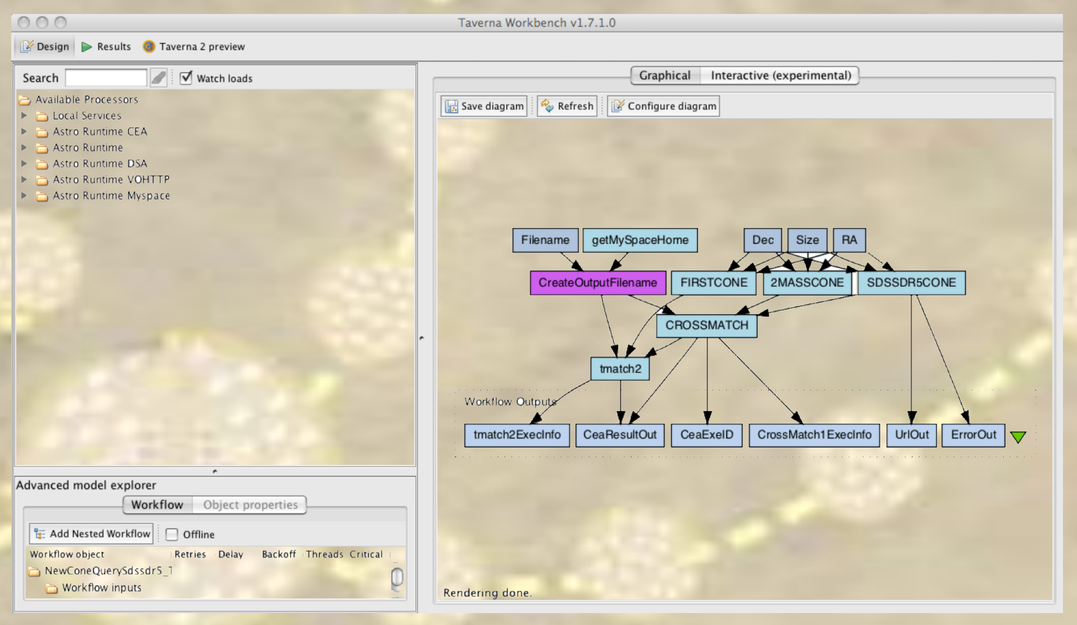

Figure 3. Astronomical workflow in the TAVERNA graphical interface. (AstroGrid. Walton & Benson 2008)

The complete digital characterization of workflows describes the scientific methodology used in an experiment in its entirety. VO services can be used as components for internet-based workflows. Since their execution is independent of the investigator's platform, they ensure the reproducibility of the results and their dissemination given their modularity, and their universal availability.

The AMIGA Project

The AMIGA project (http://amiga.iaa.es) is an international scientific collaboration led from the Instituto de Astrofísica de Andalucía – CSIC. It focuses on a multi-wavelength analysis of a statistically significant sample of isolated galaxies, in order to provide a statistical baseline to compare with the behaviour of galaxies in denser environments.

Since intensive analysis of 3D data at all wavelengths is needed for achieving the scientific goals, the group is actively involved in the migration of kinematical modeling tasks to second generation VO services deployed in a GRID environment as well as in developments of astronomical workflows and their preservation.

CSIC

INSTITUTO de ASTROFISICA de ANDALUCIA
CSIC
I·A·A

AMIGA

Analysis of the interstellar Medium of Isolated GAlaxies