

The pilot project for the TNG long-term archive

R.Smareglia (1), U.Becciani (2), A.Caproni (1), C.Gheller (3), J.C.Guerra (5), N.Lama (1), G.Longo (4), F.Pasian (1), A.Zacchei (1)
(1) INAF – Osservatorio Astronomico di Trieste, (2) INAF – Osservatorio Astrofisico di Catania, (3) CINECA Bologna, (4) Univ. Napoli "Federico II", (5) Centro "G.Galilei" La Palma

The pilot project aiming at the definition of a long-term archive for the TNG is described. Particular reference is made to the structure of the system from the hardware and software points of view, and to some possible data mining applications. Some conclusions are drawn on the requirements for interoperability with other archives, aimed at the inclusion of the TNG long-term archive in the phase-B of the European Astronomical Virtual Observatory (AVO) project.

Scope

A detailed pilot project is underway, aiming at demonstrating the feasibility of technical solutions for the Long-Term Archive (LTA) of the "Telescopio Nazionale Galileo" (TNG), within the framework of general electronic archives of astrophysical data, both existing and to be implemented in the future.

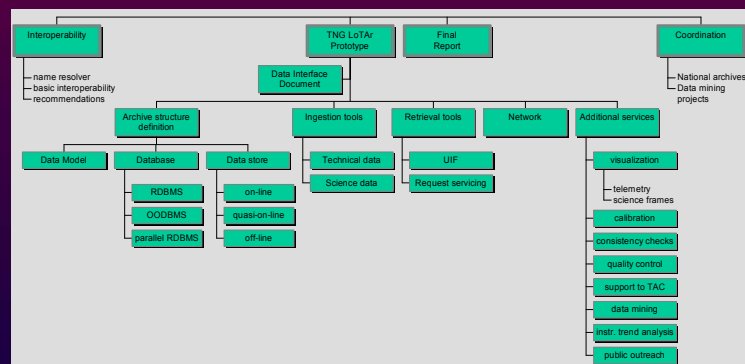
The TNG LTA will be implemented separately and at a later stage. To demonstrate its feasibility, the work plan of this pilot project includes the construction of a prototype LTA, and the study and prototype implementation of methods allowing:

- the long-term storage of scientific and technical data from the TNG;
- testing practically the feasibility of deriving value-added information from the stored data;
- guaranteeing access by the CGG staff and by the scientific community to original and derived data;
- providing tools to support the life cycle of observing proposals;
- providing interoperability with a number of existing international data centers and archives.

Furthermore, a number of suggestions are planned to be made to guarantee harmonization of the TNG Long-Term Archive with other projects related to archiving of data of astrophysical interest, with particular reference to projects specific to the Italian astronomical community, and to the national and international coordination efforts towards a multi-wavelength Virtual Astronomical Observatory. A close interaction is ensured with a number of existing national and international projects on advanced data mining tools.

Participating Institutes are: Osservatorio Astronomico di Trieste, Osservatorio Astrofisico di Catania; Centro "Galileo Galilei" - La Palma; Università di Napoli "Federico II"; Centre Données astronomiques de Strasbourg; Space Telescope European Coordinating Facility – Garching; Osservatorio Astronomico di Roma; Osservatorio Astronomico di Capodimonte; Istituto di Radio-Astronomia CNR - Bologna.

Development Plan



The diagram sketched on the left defines the development plan of the project, and is mapped to the already-described scope.

The various products of the project are the single leaves of the tree.

The key points in the development are:

- the definition of a structure for the prototype LTA, with the identification and selection of a data model, and of a proper hardware/software environment to support it;
- definition and prototype implementation of tools allowing to populate the LTA and to retrieve data from it;
- testing of the efficiency over the network, and bandwidth requirements;
- implementation of additional services (quick-look tools, on-the-fly calibration of science frames, data mining, etc.);
- implementation of a name resolver and basic interoperability tools.

A final report and a set of recommendation documents shall be issued at the end of the project.

Implementation status

Up to now, quite a number of basic issues in the implementation of the prototype TNG LTA has been tackled and solved. The data model of the TNG and consequently the data items needed to populate the archive, on the basis of the requirements defined by a dedicated WG set up in 1994, have been defined. The data interface between the archives at the telescope (AaT) and the LTA has been defined and implemented. After having designed the database structure (see aside), ingestion tools for TNG data (both technical and scientific) have been implemented, and the LTA is now being populated regularly.

As for the retrieval of data, a user interface (UIF) is being implemented on the basis of requirements defined within a group of participants in the project. A module for the servicing of user requests (including retrieval of telemetry and subsets of FITS files) has been designed and is being completed.

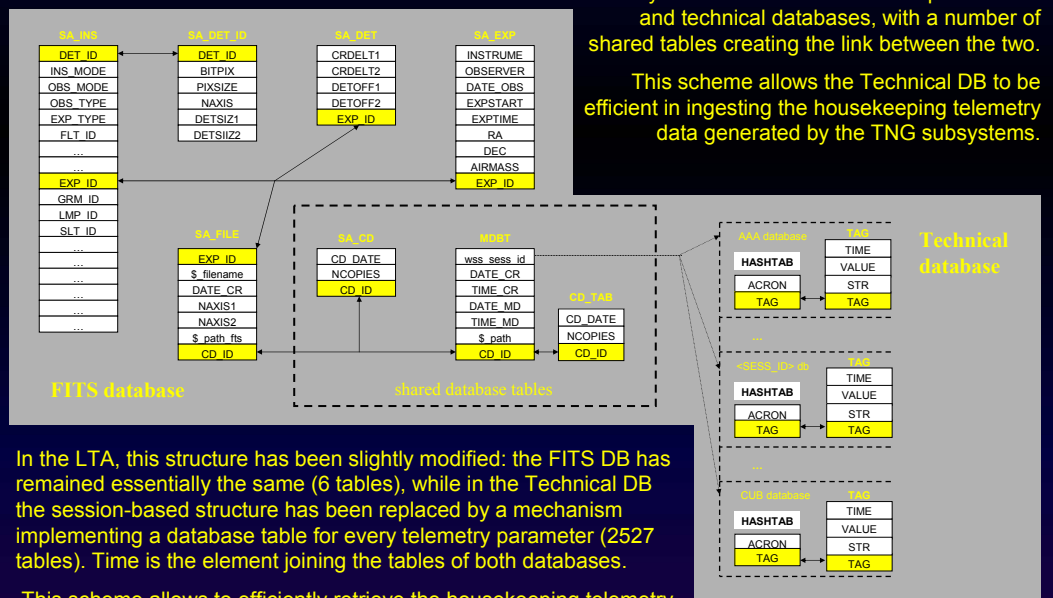
Visualization tools are being finalized as a part of the UIF. They cover both image display and X-Y graphs, offering standard options (including zooming, panning, etc.).

Benchmarking of database efficiency (see below), file retrieval efficiency for on-line, quasi-on-line and off-line data files are being carried out. Network tests will be made with a test user group when the prototype is completed.

A data mining activity is being carried out on the Technical DB. This is basically aimed at two separate goals:

- finding relations/correlations among housekeeping parameters, through the unsupervised use of neural networks;
- statistical analysis of the H/K parameters giving guiding error on the 3 axes, to analyze the mechanical behaviour of the telescope, and the possible presence of oscillations under specific conditions.

Database structure



In the figure below, the database structure of the AaT system is shown. There are separate FITS and technical databases, with a number of shared tables creating the link between the two.

This scheme allows the Technical DB to be efficient in ingesting the housekeeping telemetry data generated by the TNG subsystems.

In the LTA, this structure has been slightly modified: the FITS DB has remained essentially the same (6 tables), while in the Technical DB the session-based structure has been replaced by a mechanism implementing a database table for every telemetry parameter (2527 tables). Time is the element joining the tables of both databases.

This scheme allows to efficiently retrieve the housekeeping telemetry data from the Technical DB, while keeping the same efficiency on the science data.

Parallel access to the database

A test was made to analyze the efficiency of the access to the prototype LTA database in a parallel environment, and in particular to verify the speedup achievable with an SMP system with shared memory using the parallel query feature available in Oracle 8.1.7, the database management system chosen for the prototype LTA development. The speedup is computed as the ratio T_s/T_p (where T_s and T_p are the times required to carry out searches with 1 and 4 parallel CPUs, respectively).

Four basic tests were carried out on tables of both the Technical and FITS Databases:

- search of the FITS Database by filter name;
- single-table search on the Technical Database with predefined time interval;
- search for concurrent changes in parameters values (different Technical Database tables);
- search for out-of-range values in different Technical Database tables.

The results can be summarized as follows:

- no significant improvement is achieved in the case of small tables (case 1) or of searches made using an indexed parameter (case 2);
- a remarkable speedup has been achieved in the case of concurrent searches on different tables (case 3) with non-indexed parameters;
- in case 4, an intermediate result has been achieved: a remarkable speedup for searches on few concurrent tables (1 or 2, which is reasonably the most frequent case), negligible in the case of more tables, where the increase of the concurrent out-of-range values restricts the search range.

Interoperability in the AVO framework

One of the aims of the prototype LTA development activity is to provide tools to achieve a certain degree of interoperability with other archives at the national and international level. The following steps have been followed:

- use of internationally-defined standards in the implementation of the prototype:
 - FITS (the Flexible Image Transport System),
 - ASU (the Astronomical Server URL, a standard to generate queries for retrieving tabular data and catalogues in astronomy),
 - Astrores (a tool describing Astronomical Catalogues and Query Results with XML),
 - a preliminary version of VOTable (an extension of Astrores allowing to take binary data into account);
- installation of a name resolver using the SIMBAD facility;
- definition of a higher level of interoperability based on the availability at the CDS of the TNG catalog of observations.

The link with international projects is actively pursued by participation of the TNG LTA group to the Science WG of the AVO (a project, funded by the EU to provide a prototype multiwavelength virtual astronomical observatory for the scientific community) and to the Interoperability WG of the OPTICON project. The latter is a more technical group where standards are defined, discussed and agreed upon; and where the major data providers and of the virtual observatories projects (AVO, NVO/USA, AstroGRID/UK) are represented.

This international activity is felt to be particularly important, in view of the possible creation of a national astro-grid for computation and data access, and of the eventual participation of the Italian astrophysical community in the phase-B of the AVO project.

The proposal for the implementation of this pilot project has been submitted as a response to a CNAA Announcement of Opportunity (Bando n. 5/2000) and has been completely funded. The completion date is planned to be end of June 2002.

