RevMexAA (Serie de Conferencias), **42**, 86–87 (2013)

THE GTC SCIENTIFIC DATA ARCHIVE

E. Solano[1,2] and R. Gutiérrez[1,2]

RESUMEN

El archivo del Gran Telescopio CANARIAS es el resultado de un acuerdo de colaboración entre INTA y GRANTECAN. Desarrollado en el marco del Observatorio Virtual Español, el sistema es gestionado por la Unidad de Archivos del Centro de Astrobiología (INTA-CSIC). El archivo se ha diseñado siguiendo los estándares definidos por la Alianza del Observatorio Virtual Internacional (IVOA) lo que garantiza el alto nivel de interoperabilidad que se espera de un telescopio de la relevancia de GTC. En este artículo se describen las principales características y funcionalidades del archivo de datos de GTC.

ABSTRACT

The GTC Archive is the result of a collaboration agreement between INTA and GRANTECAN. It has been developed in the framework of the Spanish Virtual Observatory project and is maintained by the Data Archive Unit of the Centro de Astrobiología (INTA-CSIC). The archive has been designed in compliance with the standards defined by the International Virtual Observatory Alliance which will guarantee the high levels of data accessibility and handling demanded by a telescope of the significance of GTC. In this paper we describe the main characteristics and functionalities of the GTC Scientific Archive.

Key Words: astronomical databases: miscellaneous — virtual observatory tools

1. INTRODUCTION

In the last years, there is an increasing awareness by scientists and funding agencies of the need to maximize scientific return on the costly investments required by current astronomical projects. In this context, astronomical archives play a fundamental role ensuring this return as demonstrated by the intensive usage the scientific community makes of these facilities. The long-standing IUE (with every single spectrum retrieved more than five times from the archive), and HST (with more papers published with archived data than with newly acquired data) projects are paradigmatic examples of efficiency in scientific data exploitation. This, together with the fact that archival information is dramatically growing in size, has made of archives a fundamental research infrastructure for modern astrophysics.

Nevertheless, the usability of astronomical archives has been, until recently, limited by an important factor: the lack of standardization (different access and retrieval protocols, data models, formats,...) among the astronomical archives and services makes it very inefficient the data discovery, query, gathering and analysis from more than one resource.

The Virtual Observatory[3] (VO) is the international initiative that attempts to solve this problem. The VO concept goes one step further than just giving access to distributed computational resources or to the data. It also permits operations on the data and returns results. VO is opening new lines in the astronomical research as demonstrated by the growing number of VO-refereed papers. Spain takes part in this endeavour since 2004 through the Spanish Virtual Observatory[4] (SVO) whose core team is located in the Data Archive Unit[5] of the Centro de Astrobiología.

Having an archive perfectly integrated in the VO framework constitutes an added value of enormous importance for an astronomical project. This was clearly understood, already in the early stages of the project, by the GTC responsibles who selected the SVO group at CAB for the development and maintenance of the VO-compliant GTC Scientific Data Archive.

2. THE ARCHIVE: CHARACTERISTICS AND FUNCTIONALITIES

2.1. *Transfer, ingestion and storage*

Once the proprietary time is over, the data are transferred through a secure connection from GTC

[1]Centro de Astrobiología (INTA-CSIC), Departamento de Astrofísica, PO Box 78, 28691, Villanueva de la Cañada, Madrid, Spain (esm,raul@cab.inta-csic.es).
[2]Spanish Virtual Observatory, Spain.

[3]`http://ivoa.net`.
[4]`http://svo.cab.inta-csic.es`.
[5]`http://sdc.cab.inta-csic.es`.

to CAB where a number of control tests are carried out to guarantee a high level of quality and coherence in the metadata. Problematic files are inspected and remedial actions are proposed in agreement with the GTC staff. After passing the tests, the metadata information is extracted from the FITS headers and inserted in the database system whereas the FITS files are moved into the data storage system. The database and the web servers are hosted in two different HP Proliant 8-processor systems. Two other HP Proliant 4-processor servers are devoted to development activities and a dedicated PC to the data transfer between GTC and CAB. The servers are connected through optical fiber to the storage unit, an HP EVA-4000 disk array with 20 TB capacity. To ensure data integrity, a well-defined backup and data safekeeping policy has been established.

2.2. *Query form*

The friendliness of the archive user interface is extremely important for the archive itself to be effectively used by the community at large. The user interface is a web-based fill-in form that permits queries by a number of parameters, namely: lists of objects/coordinates, observing date, instrumental configuration and observing program. The system incorporates a name resolver allowing queries by any of the names provided by the SIMBAD[6] database. The output fields may be ordered by date, instrument or program. The number of results shown per page can also be customized.

2.3. *Results from search*

Apart from the list of results that meet the criteria entered in the query, the system implements a number of functionalities to allow an effective use of the archive according to the user needs. Some examples are:

• Multidownload capabilities: In the table of results, each raw scientific file is associated with its corresponding calibration, acquisition and quality control files. This permits users an easy identification and retrieval of the necessary datasets to perform the data reduction. Data can be retrieved individually or in groups from the results page. For multiple-file retrieval it is possible to include or exclude individual datasets. Multiple-file download generates a file in ZIP format.

• Visualization tools: Data visualization is carried out using Aladin[7], a VO-compliant software that allows users to superimpose entries from astronomical catalogues or data bases available from VO services. Also, FITS headers can be visualized by clicking on the corresponding link.

• Access to documentation and HelpDesk: Links to web pages providing detailed description of the GTC project as well as of the archive system are available. A HelpDesk system allowing users to directly contact the archive staff is also implemented. A *Frequently Asked Question* section with the answers to the most common questions posed by the users will be implemented in subsequent versions.

3. INTEROPERABILITY: THE VIRTUAL OBSERVATORY

It is a fact that the linkage across archives and wavelengths regimes in the framework of the Virtual Observatory initiative adds still more value to the GTC data as makes possible research projects that would be extremely difficult to perform outside VO. A major drawback in this sense is the lack of an official reduction pipeline for GTC data (something that will be partly solved for OSIRIS in the near future) since VO expects fully-reduced, science-ready data. In a first stage, and once these science-ready datasets are available, GTC will implement three types of VO services: SIAP for astrometrically (and ideally photometrically) corrected images, SSAP for reduced spectra and ConeSearch for high level products like catalogues. In all cases, a client will search through the GTC VO services for available data that match certain client-specified criteria using a HTTP GET request. The response is a table (in VOTable format) describing the available data, including metadata and access references (implemented as URLs) for retrieving them.

4. CONCLUSIONS

A well-designed, properly-implemented, VO-compliant scientific data archive, as part of the capabilities of the GTC telescope, is a major contribution toward the full exploitation of its unique characteristics. The access to GTC data by the wider community after the proprietary period is over will ensure that the information are fully exploited and that maximum scientific value is returned to the astronomical community and, ultimately, the citizens that support GTC through taxes. An astronomical archive is more than the sum of the parts acting alone and, therefore, the collective impact of the GTC archive will far exceed what could be produced by the programs of individual observers or teams.

[6] `http://simbad.u-strasbg.fr/simbad/`.
[7] `http://aladin.u-strasbg.fr/`.