*Highlights of Spanish Astrophysics VII, Proceedings of the X Scientific Meeting of the Spanish Astronomical Society held on July 9 - 13, 2012, in Valencia, Spain. J. C. Guirado, L.M. Lara, V. Quilis, and J. Gorgas (eds.)*

# The Gran Telescopio Canarias and Calar Alto public archives.

**E. Solano[1,2], R. Gutiérrez[1,2], J.M. Alacid[1,2], F. Jiménez-Esteban [1,2,3], and A. Velasco[1,2]**

[1] Depto. Astrofísica, Centro de Astrobiología (INTA-CSIC), P.O. Box 78, E-28691, Villanueva de la Cañada, Madrid, Spain;
[2] Spanish Virtual Observatory
[3] Sant Louis University, Madrid Campus, Division of Science and Engineering , Avenida del Valle 34, E-28003, Madrid, Spain

## Abstract

In this paper we describe the main characteristics and functionalities of the Gran Telescopio Canarias (GTC) and Calar Alto public archives. The archives have been developed in the framework of the Spanish Virtual Observatory and are maintained by the Data Archive Unit at Centro de Astrobiología (CAB, INTA-CSIC). The archives contain both raw and science ready data and have been designed in compliance with the standards defined by the International Virtual Observatory Alliance (IVOA) which guarantees a high level of data accessibility and handling.

## 1 Introduction

Astronomical infrastructures are expensive and scientists and funding bodies have the obligation in front of taxpayers to maximize the scientific return of these costly investments. Although archives are key pieces to guarantee the optimal exploitation of astronomical datasets, the potential of the archive-based research has been, until recently, limited by an important factor: The lack of interoperability and standardization which makes it very inefficient the data discovery, query, gathering and analysis from more than one archive.

The Virtual Observatory (VO) is the international initiative that attempts to solve these problems. VO is opening new lines in the astronomical research as demonstrated by the growing number of VO-refereed papers. Spain takes part in the VO initiative since 2004 through the Spanish Virtual Observatory (SVO) whose core team is located in the Data Archive Unit at CAB.

The Gran Telescopio Canarias (GTC) is the largest telescope in the world (10.4m segmented primary mirror) in the optical-infrared regime. It is located in the Observatorio del Roque de los Muchachos (La Palma, Spain) and started scientific operations in the Spring of 2009. GTC has two Day One instruments: OSIRIS (an optical camera and multi-object spectrograph) and CANARICAM, a thermal infrared camera and spectrograph with polarimetry and coronography capabilities. The Calar Alto Observatory is the largest astronomical observatory in continental Europe. It presently consists of three telescopes (1.23m, 2.2m, 3.5m) with a suite of imagers and spectrographs operating in the optical and infrared regime.

Having an archive perfectly integrated in the VO framework constitutes an added value of enormous importance for an astronomical project. This was clearly understood by the GTC and Calar Alto board committees who selected the Spanish VO group at CAB for the design, development, implementation and maintenance of their VO-compliant data archives.

## 2 The archives

### 2.1 Overall characteristics

The database and the web servers of the GTC and Calar Alto archives 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 the observatories 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 defined. To guarantee an efficient usage of the existing resources, the GTC and Calar Alto archives have been developed using the same technologies, namely, Java as programming language and Postgress as database engine.

Both GTC and Calar Alto are evolving infrastructures. This requirement has been taken into account in the design of archive system, which is modular enough to easily accommodate future instruments and telescopes as well as new functionalities.

### 2.2 Data transfer and ingestion

Once the proprietary time is over, data is transferred through a secure connection from GTC and Calar Alto to CAB. At their arrival, a number of control tests are performed to assure data integrity and metadata coherence. Problematic files are inspected and remedial actions are suggested in agreement with GTC and Calar Alto staff. After passing the tests, the metadata information is extracted from the FITS headers and stored in a database whereas the FITS files are moved into the data storage system. Once the data and metadata have been successfully ingested, they are automatically available to the general public through the web and VO interfaces.

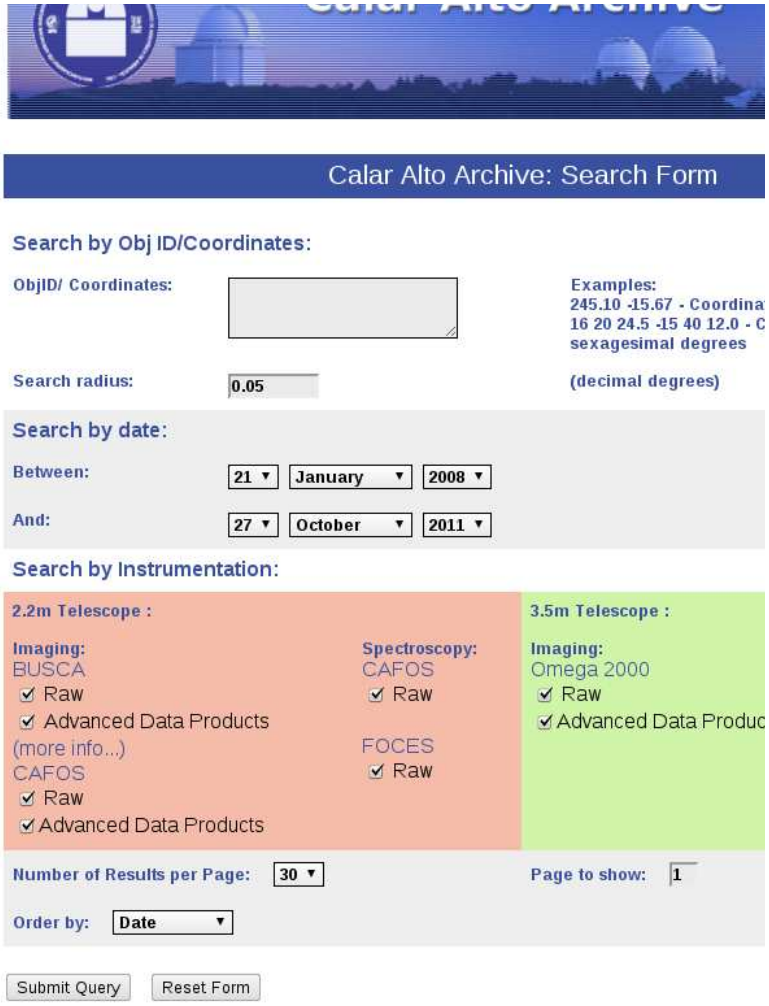


Figure 1: Calar Alto archive. User interface.

## 2.3 Web interfaces

Both the GTC and Calar Alto web interfaces are friendly enough to be potentially used by a wide variety of users, ranging from curious people to professional astronomers. The archives can be consulted by typical parameters (coordinates and radius, observing time, instrumental configuration, ...). The output fields can be sorted and the number of results shown per page can also be customized (Fig. 1).

In the table of results, each raw scientific file has associated its corresponding calibration files which permits an easy identification and retrieval of the necessary datasets to perform an off-line data reduction.

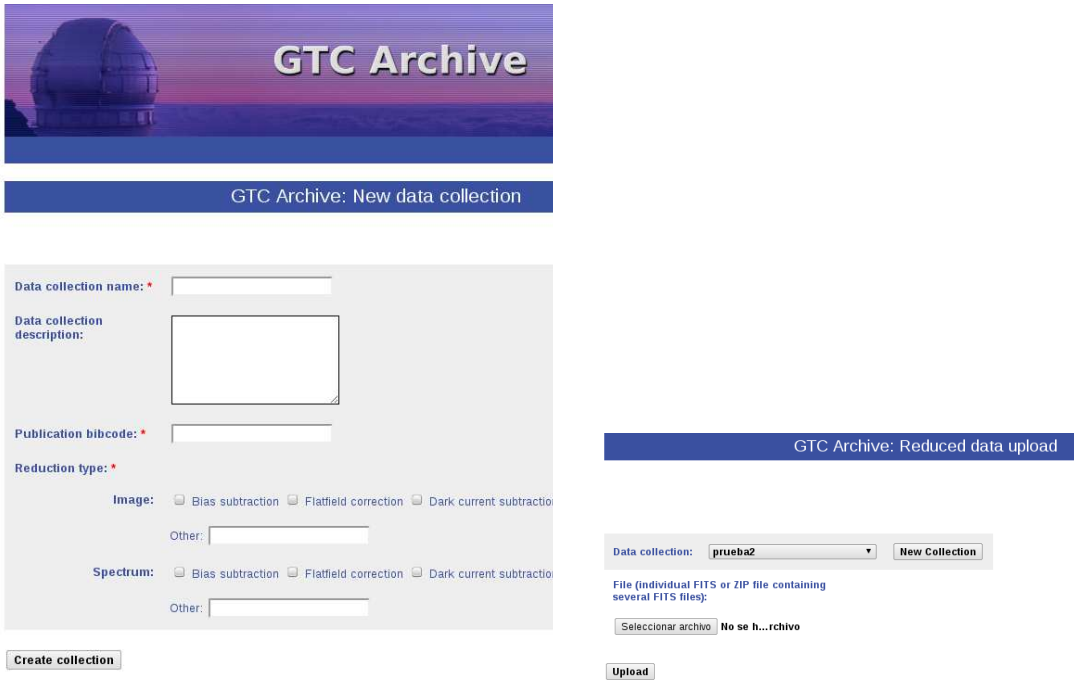


Figure 2: Schematic view of the application to ingest community HLDPs into the GTC archive.

# 3 High Level Data Products (HLDPs)

High Level Data Products (photometrically and astrometrically corrected images, spectra ready for immediate scientific exploitation, catalogues, mosaics, stacked images,...) are of fundamental importance for archives as they enhance their use by the community. Moreover, high level data products provide a higher visibility of the project results as clearly demonstrated by the large number of refereed papers based on archived data from projects like SDSS, 2MASS, UKIDSS and WISE, among others.

## 3.1 HLDPs at the GTC archive

Like other astronomical archives (e.g. ESO), we have developed a system for the community to return GTC reduced data to the archive. The procedure is as follows:

Firstly, CAB staff looks for refereed publications (ApJ, AJ, A&A, and MNRAS) containing GTC data. Then, the first author of each of these publications is contacted and invited to send us the reduced data used in the paper. To ensure the successful integration of these data products and in order to carry out the data ingestion in a way as automated and transparent for the user as possible, an ingestion application has been developed (Fig. 2). This application performs the validation process in two steps:

- Matching of the reduced products with their corresponding raw products. This matching is made automatically using specific keywords that must be present in the reduced product.
- Data verification. The reduced data are analyzed to ensure integrity, coherence and uniqueness in the archive. If the reduced product has been already uploaded, this is detected and reported. The existence of compulsory keywords needed to ensure the compliance with the Virtual Observatory data models is also tested. Once the verification process has been successfully, the corresponding results are accesible from the archive.

### 3.2 HLDPs at the Calar Alto archive

Data Centres play a fundamental role in the provision of High Level Data Products. In this sense, we have started to build two types of HLDPs in the Calar Alto archive:

- Astrometrically corrected images: Calar Alto raw images do not include the World Coordinate System (WCS). Using the software Astrometry.net we have added astrometric metadata which allow the transformation between image and sky coordinates. The quality of the astrometric correction is assessed using Sextractor which computes the positions of the sources in the images, sources that are compared with counterparts in the 2MASS Point Source and USNO-B1.0 catalogues. Metadata containing quality information (number of Sextractor sources in common with USNO-B1.0 and 2MASS, mean, standard deviation and median of the differences in position) are included as additional keywords in the astrometrically corrected images.
- Source catalogues, extracted from Calar Alto astrometrically corrected images. In particular, we are building a catalogue of asteroids observed serendipitously or as primary targets in CAFOS images (Fig. 3). The identification of the asteroids is done using the SkyBot VO service. Once confirmed, the asteroid coordinates are accurately determined an sent, together with the temporal information, to the Minor Planet Center which uses it to refine the associated orbital parameters.

## 4 The Virtual Observatory services

The Virtual Observatory is the framework which guarantees an efficient linkage among astronomical archives. The implementation of VO services in the GTC and Calar Alto archives represents an added value of fundamental importance and facilitates research projects that would be extremely difficult to perform outside VO. Three types of Virtual Observatory services are being implemented for the science-ready data described in the previous section: SIAP, for astrometrically and photometrically corrected images, SSAP for reduced spectra and ConeSearch for catalogues. In all cases, a client will search through the GTC and Calar Alto VO services for available data that match certain client-specified criteria using a HTTP

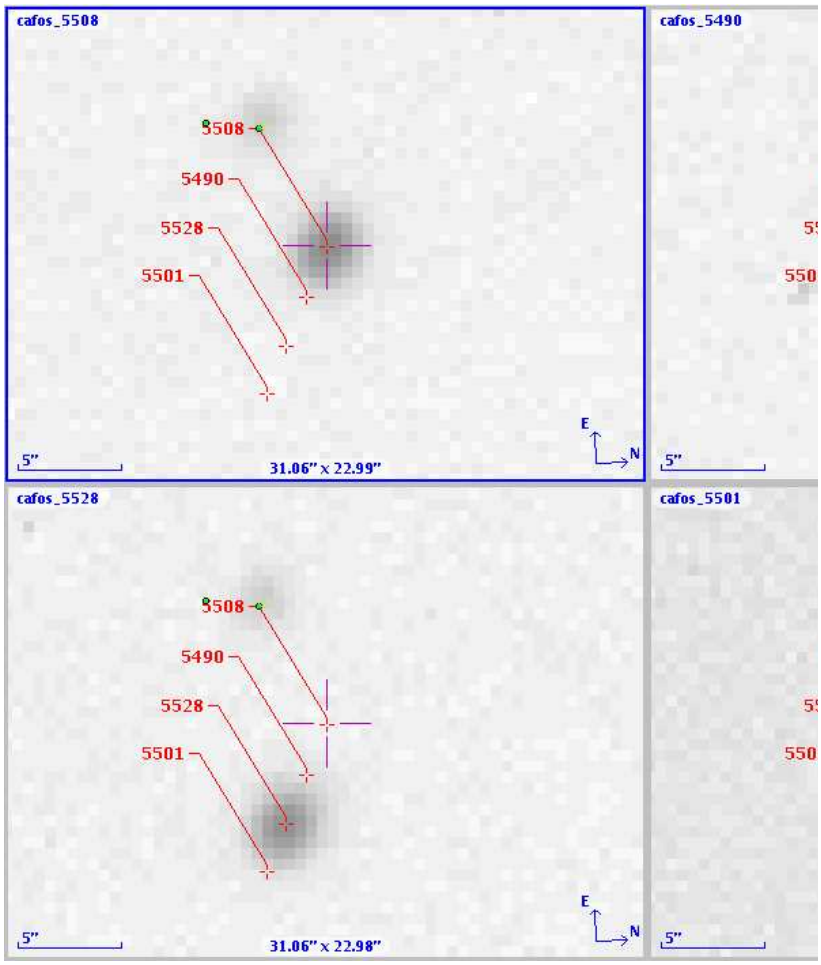


Figure 3: Asteroid 38607 observed in a sequence of CAFOS images.

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.

## Acknowledgments

This research has made use of the Spanish Virtual Observatory (http://svo.cab.inta-csic.es) supported from the Spanish MICINN / MINECO through grants AyA2008-02156, AyA2011-24052.Partially funded by the Spanish MICINN under the Consolider-Ingenio 2010 Program grant CSD2006-00070: First Science with the GTC (http://www.iac.es/consolider-ingenio-gtc)