Astronomical Data Analysis Software and Systems XXVI
ASP Conference Series, Vol. 521
Marco Molinaro, Keith Shortridge, and Fabio Pasian, eds.

Asiago Astronomical Archive: Status and Features

Elisa Londero,[1] Sonia Zorba,[1] Marco Molinaro,[1] Cristina Knapic,[1]
Lina Tomasella,[2] Aldo Frigo,[2] and Riccardo Smareglia[1]

[1]*INAF - Trieste Astronomical Observatory (OATs), 34143 Trieste, Italy;*
londero@oats.inaf.it

[2]*INAF - Padova Astronomical Observatory (OAPd), 35122 Padova, Italy*

Abstract. We present the new archiving system that has been installed at INAF OAPd - Asiago Observatory. Its main features comprise fast and reliable distribution capabilities, prompt monitoring and easy access to the data through a newly-designed web-interface.

1. Introduction

Geographically distributed archives present challenges when ingesting and delivering large amounts of astronomical data. The main issues are related to remote control and configuration, monitoring and logging anomalous conditions, fault tolerance and error handling. The New Archiving Distributed InfrastructuRe (NADIR, Knapic et al. (2014)) developed and implemented within the Italian Astronomical Archive (IA2) project, has been shown to successfully overcome these issues. NADIR has been tested extensively, it is currently installed and working at the Large Binocular Telescope Observatory (LBTO) and at Telescopio Nazionale Galileo (TNG) and we are in the process of extending its use to manage the data coming from other facilities. In this work we describe the features of the new archiving system that has been installed at INAF OAPd - Asiago Observatory. Its main components are a distribution and monitoring system based on NADIR, a new web-interface for accessing the data and a temporary archive at Asiago observatory for fast retrieval of the most recent astronomical data.

2. The project

The Italian Astronomical Archive (IA2) is a research infrastructure project that coordinates different Italian national initiatives for the improvement of astrophysical data service quality. In particular, it aims at implementing a strategy for the preservation, curation and access of astronomical data. IA2 is currently ingesting data from TNG, LBT and Asiago observatories. For the future, besides Asiago, there are plans to incorporate in the NADIR-based IA2 archiving constellation also Serra la Nave (optical), Medicina (radio) and Noto (radio) telescopes. The archive hosts also data from surveys (both raw and calibrated), simulations (ITVO) and EU projects (for example INDIGO-DataCloud, GENIUS, VIALACTEA) and it is in charge of data publication in the Virtual Observatory (VO).

Asiago Observatory was founded in 1942 and it is operated jointly by the National Institute for Astrophysics located in Padova (INAF-OAPd) and the department of Physics and Astronomy at Padova University. It hosts the largest optical telescope in Italy. The observatory is spread over two locations: the main station (at 1050 m), which hosts the Galilean telescope (122 cm) belonging to Padova University and cima Ekar (at 1366 m), which hosts the Copernicus (182 cm) and the Schmidt (67/92) telescopes, both belonging to INAF-OAPd. The incoming data rate is about 300 GB/year and the observatory is used both for research and educational purposes.

There are challenges when building archiving software, especially in connection to data evolution. The main issues comprise changes in data format, variations in publication policy and in metadata content. Software that is flexible and configurable can overcome these problems, moreover, it can be reused in different contexts, it is easily deployable and has scalability potential. Another sought-after feature of data distribution software is the capability to operate the archiving system remotely, through efficient monitoring, event logging and error tracking. NADIR was planned and built in order to take all these requirements into account. It is based on TAco Next Generation Object (TANGO, `www.tango-controls.org`) from which it inherits its main points of strength.

TANGO is a free, open source, device-oriented controls toolkit designed for managing remotely any kind of hardware or software. It is essentially a control system in which devices are managed and monitored over a distributed network and it is customarily used to control physics experiments. A service can be directly associated and implemented in a device, which is the fundamental element in TANGO. This offers a convenient way to uniformly interact with all the devices, hiding network complexity and device inhomogeneity. TANGO supports bindings and full support for three different languages (C++, Java and Python).

Building an archiving software based on TANGO provides automatically the standardization of logging and configuration retrieval activities, high scalability, modularity and robustness. In our context, a TANGO device is an archiving software application. The devices fulfill a number of tasks, from selecting the correct data format and checking its consistency to transferring the data and metadata over the network to the archiving sites. Device properties are modifiable at any time and a device once implemented, can be reused for different archiving purposes. The flexibility deriving from these conditions gives the basis for a robust and reliable service. Furthermore, TANGO distributed control system is responsible for services start up, shut down and keep alive processes. Besides the multi language environment, the final user can benefit from the complexity of communication between remote servers being hidden.

3. System configuration

In the remaining part of the paper we will detail the constituents of the NADIR archiving system in use for INAF OAPd - Asiago Observatory. The main agents ruling the archive are the TANGO servers. They act in a concerted way from the very moment the data is collected from a telescope in Asiago to its final placement in the archive in Trieste. More in detail, the data coming from an observing session is placed in a dedicated folder where an initial selection is done by a TANGO device called *preProcessor*. The conformity of the files to the FITS standard is tested and in case of a positive outcome, the files are copied to a second folder. A second TANGO server acts at this point:

the *fitsImporter*. This device extracts the metadata from the FITS cards of the selected files according to the keys stored in the datamodel and fills the instrument tables of a MySQL database. It also moves the files to the archive. Finally, there are two TANGO servers (*dataExporter* and *metadataExporter*) waiting for connections from the Trieste IA2 archiving facility. In Trieste, in order to complete the transfer process, two more TANGO servers are needed, namely the *dataImporter* and the *metadataImporter*. The first connects to the *metadataExporter* and asks for an update of the metadata based on the last registered timestamp. The second, after authenticating on the *dataExporter*, asks for the download of the list of files obtained by the *metadataImporter*. A list of files whose download has failed is also updated regularly by the *dataImporter*. The main archive hosted in Trieste is available for public access compliant with the publication policy. Figure 1 describes the details of the scheme used for the archiving procedure.

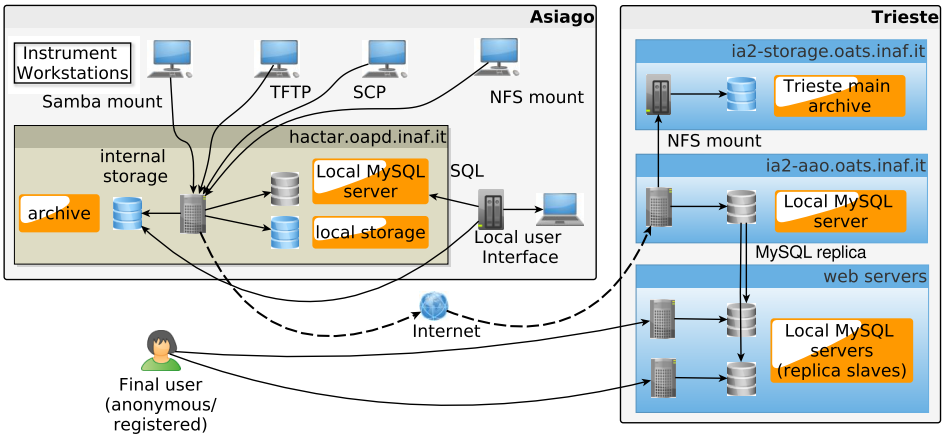

Figure 1. Sketch of the archiving system used for processing the data coming from the Asiago observatory. Data from observations are stored in the archive and metadata used for data retrieval are inserted in the MySQL database. The user can locally access the data through a user interface. After being transferred over the Internet, data and metadata reach the master machine in Trieste. The data is stored in its final destination while the metadata database undergoes a double replication on two different machines which grant access to astronomical data both to the anonymous and the registered user.

A temporary archive with data less than three months old is available from within the LAN at the observatory. This additional service is meant to reduce latency issues in case data retrieval of the most recent images is needed. In order to speed-up and systematize the archiving process, a tool that makes it possible to amend FITS file values in case of an accidental incorrect insertion has been deployed at Asiago observatory. The file can be amended directly by the astronomer and the correction is propagated up to the main archive in Trieste. This practice has also the advantage of reducing the maintenance burden on the archivists.

It has been shown that NADIR makes the archiving process more systematic and less error-prone and for this reason there are plans to further extend its use. At the same time there is the need to systematize the way the final user has access to the archived data. To this end, the APOGEO (Automatic POrtal GEneratOr) web-interface

generator (Zorba et al. 2016) has been developed within the IA2 project. Its main goal is to generate portals for access to astronomical archives in an automatic, standardized and reliable way. The portal administrator is only required to configure the portal through a graphical wizard; subsequently the application itself generates a package that is ready to be deployed on an application server. The access to the resources is done via queries on the main instrumental and observing parameters and the query result is exportable via SAMP protocol to the Virtual Observatory compliant clients like Aladin or Topcat. As for now the authentication is managed through the IDEM (IDEntity Management) federation but there are plans to employ eduGAIN in the future. A web interface based on APOGEO has been deployed to access the archive locally in Asiago and publicly in Trieste.

4. Conclusions

The archiving system detailed here has been installed and tested at INAF-OAPd, Asiago Observatory. Besides the advantages introduced by the NADIR archiving software in terms of flexibility towards data evolution and easy monitoring of the archiving steps, the deployment provides also a completely new web interface for accessing the data. While the main public archive is hosted in Trieste, a temporary archive accessible only from within observatory's LAN in Asiago reduces latency issues in the case of data retrieval. The system described here makes the data transfer from INAF OAPd - Asiago Observatory to Trieste archiving facility more systematic and organized, reducing the chances of error and facilitating the monitoring procedure through all the archiving steps.

References

Knapic, C., et al. 2014, in ADASS XXIII, edited by N. Manset, & P. Forshay, vol. 485 of ASP Conf. Ser., 131

Zorba, S., et al. 2016, Proc. SPIE, 9913, 991344. URL `http://dx.doi.org/10.1117/12.2232594`