ВАК 2024
MODERN ASTRONOMY:
FROM THE EARLY UNIVERSE TO EXOPLANETS AND BLACK HOLES
300 years Russian Academy of Sciences

Information system and general archive of SAO RAS observations

O. Zhelenkova, T. Plyaskina, V. Shergin, G. Mal'kova, A. Velichko, and M. Kyabishev

Special Astrophysical Observatory of the Russian Academy of Sciences, Nizhny Arkhyz, 369167 Russia

Abstract. The telescopes of the Special Astrophysical Observatory of the Russian Academy of Sciences provide unique observational material. For observations, various astronomical instruments and data acquisition systems are used, which record observations with different data structures and descriptive parameters. The general archive of the observatory contains more than two dozen digital collections with observations obtained on various instruments. Maintaining an archival system requires software development, metadata management, fault tolerance, and periodic migration of files to modern storage media, which is important to ensure the long-term preservation of digital information. The archival system consists of three interconnected components that provide accumulation, long-term storage, and access to the data. The information system is implemented based on free PostgreSQL. The access to the archived files is provided through a web interface that supports queries by device, date, coordinates, object name, program applicant, and observer. The archive size is about 3 TB and its database contains more than 4 million records. The archive system uses two servers with directly connected storage systems. The main server implements data access, and the test server is used for development and testing. This configuration provides both permanent access to the files and the ability to modernize the information system.

Keywords: methods: observational, data analysis; astronomical databases: miscellaneous

DOI: 10.26119/VAK2024.184

1 Introduction

The history of the archive of SAO RAS observations began in the early 80s of the last century, when the accumulation of the data obtained on RATAN-600 and BTA had started to be carried out in digital form. This was the impetus for the development of the concept of a digital data bank (Vitkovskij et al. 1987; Kononov et al. 1990) as well as for the work on standardizing the observation file formats. The first digital archive was organized for the surveys of the Cold Experiment (Berlin et al. 1984). The self-documenting FITS format (Wells et al. 1981), which is the de facto standard for the presentation of astronomical data, came into use at the observatory in the late 1980s. This standard was first adapted to our observations and applied to the files (Vitkovskij et al. 1988) obtained with the observatory's first CCD camera (Borisenko et al. 1990). The FLEX format similar to FITS was then developed for the continuum radiometer data and used for the RATAN-600 observations.

As the data volume grew, the moment came when the information from the reel-to-reel magnetic tapes had to be rewritten onto various digital media such as streamer tapes, data cassettes, magneto-optical disks, and, finally, onto optical disks, which had begun to be used for permanent storage of the archive data since 1994. The experience of migrating the information and storing the data on streamers and magneto-optical disks was not very successful, since some part of the observations were lost.

According to the concept of an observatory databank (Vitkovskij et al. 1987), the general archive consists of local archives, with the local archive being a digital collection obtained by a single observing method provided by an astronomical instrument and its data acquisition system. The data in the general archive is stored in FITS and FLEX formats, the internal file format of the MIDAS system (Warmels 1992), as well as in binary files with text descriptions archived in tar files.

By 1999, the archive had contained more than 100 GB of data, and a search system had already been needed to access the archived files. Together with colleagues from the South Russian Regional Center for Informatization of Rostov State University and based on the local archive of the radiometer data from RATAN-600 feed cabin No. 1, a prototype of an information system with web access based on the Oracle DBMS[1] was created (Vitkovskij et al. 2000). For each observation in the database, a service block was formed from the parameters of the observation file recorded in the file header. It also included the attributes to identify the file.

In further work with the local archives, it turned out that the service information block contains an excessive number of parameters. We had to limit ourselves to

[1] https://www.oracle.com/cis/

those attributes that had been present in the files of all the local archives. And this attribute turned out to be the date of observation. Taking this into account, as well as the experience of the developing of the prototype, in 2003 we implemented an information system based on a trial version of Oracle. The information system had implemented the search by observing date in 14 local archives with the ability to copy files via a web interface (Zhelenkova et al. 2003, 2004).

To expand the number of the types of queries to observation files, a next version of the search system based on PostgreSQL[2] was implemented. It had provided data search using the standard set of queries. To describe each file in the database, the number of attributes had been increased, including the identification and location of the file in the storage space. In 2003, the IAU General Assembly passed a resolution[3] on the observations obtained by the publicly funded observatories, which noted that an observatory was obliged to place data in the public domain after the copyright of the applicants for the program expires. These provisions were taken into account in a new version of the search system (Zhelenkova et al. 2007, 2010).

Since 2013, using the software developed by Shergin (2014) we have been performing automatic coordinate calibration of direct images and/or correction of file headers. Files processed in this way are written to special local archives. To date, more than 300 thousand files from 4 local archives have been processed in this way.

In 2018, radio data was added that had to be migrated to our system from another archive system that was no longer supported. This doubled the number of files in the information system. At first, we implemented the access to files based only on observation date. Since 2023, the search for radio observations has been carried out using the standard set of requests.

The archival system is regularly updated with new data. The local archives of decommissioned instruments continue to be maintained. When new observing instruments are put into operation, a new local archive is organized and added to the system.

Basic information about the current state of the observatory's general archive of observations is presented in Zhelenkova et al. (2017, 2019, 2023).

2 The archive system

The general archive stores the data obtained with the observing methods used on the observatory's telescopes. This data includes observations, service files used in

[2] https://www.postgresql.org
[3] https://www.atnf.csiro.au/people/Ray.Norris/WGAD/Resolution.htm

processing the raw data, and observation logs. The archival system consists of three components:

(1) accumulation—cascade scheme of archiving;
(2) storage—hardware and software for permanent data storage;
(3) search and access—information system with web access.

The archive system has two databases located on two servers. Each server storage space has same structure and content. One server supports the working version of the archive, the second one supports the test version, with which we carry out and test all new developments. The support of the two databases ensures consistent data addition.

2.1 Cascade archiving scheme and storage space

The cascade archiving scheme defines the flow of data from the acquisition system to permanent archival storage. First, during observations, the data is accumulated on the acquisition system computer of this observing method. When the observing program changes, the data is moved to a dedicated space on the file server intended to the observing method. Then, from the observations obtained by one method, the image of an archive disk is formed. Its structure is formed according to the following rules:

(1) a label is generated with the disk number and the name of the local archive;
(2) observations from one night are placed in one directory, the name of which corresponds to the date of the observations;
(3) one observation is recorded in one file.

Then the image is written to an optical disc. Such a disk is part of the permanent offline storage of the archival system.

To ensure the integrity of data in case of destruction of physical media or input/output errors, we have two levels of the archive storage. The offline level is closed to user access. It includes two copies of optical disks with archival data and is used by an archive administrator when they need to restore data in the case of emergency failures.

The archive storage space on a dedicated server includes the PRIME directory, in which the copies of the optical disks are located, and the DATA directory, in which the copies of the disks are placed after verification and correction by our software tools.

Each disk is copied into the PRIME and DATA spaces into directories with names corresponding to the disk number in the offline level of the archive. The physical structure of the directories is superimposed on the logical structure, reflecting the distribution of the disks among local archives.

2.2 Information system

The information system of the general archive of observations is implemented in the PostgreSQL DBMS. Each observation file is described in the database tables by about 30 attributes. They are used for the dynamic formation of the web interface, for mapping the FITS parameters, and for file identification.

The database schema includes two dozen tables and views. The tables of the information system can be divided into three groups based on the frequency of adding new records. The first group includes the tables which are filled in when creating a database. New records can only be added in these tables when a new local archive is added into the information system. Such uploads do not occur often, so these tables can be considered as static. The second group comprises the tables that can be updated during CD/DVD disk analysis when a new item is added into the system for the first time, for example, a new observing program, an observer, etc. The third group includes the tables with the information about each archived file. Entries are added to the tables when ingesting new disks. In this process the tables of the first two groups are used for the analysis of a new disk. A special place is occupied by the table that links the attributes of the tables which store the information about observation files, with FITS header keywords and UCDs (Unified Content Descriptors) (Derriere et al. 2004).

3 Conclusion

Since astronomical observations do not lose scientific significance over time, their preservation is one of the main tasks of a digital archive.

Permanent archival storage means storing digital data collections for decades. This requires the mandatory storage, along with the collection files, of the description of its organization and metadata that allows the information to be interpreted. A permanent archive must search, access, and display the digital objects it stores, even if storage technologies change.

Since 2003, we have been taking into account these circumstances when developing and maintaining the archival system of the observatory. This experience of the management of various digital collections allowed us to develop the following approach to data preservation: data duplication, and on various types of digital media. To ensure the durability of the information system, the duplication of the database have been implemented on two dedicated servers. If one server fails, the second server can provide access to data. This also allows the archival system to be developed both in terms of software and in terms of adding new collections, because these actions are performed on the test server, and after verification they are transferred to the working server. We also note that this organization of the archival system allows us to rewrite digital collections onto a new type of media.

Taking into account the existing equipment, the prospect of increasing data volumes, and modern technologies for supporting the persistent archives, we are considering a possibility to organize software-defined storage for our archive based on the iRODS software (http://irods.org) or similar systems (Zhelenkova et al. 2020).

References

Berlin A.B., Gassanov L.G., Gol'Nev V.Y., et al., 1984, Soobshcheniya Spetsial'noj Astrofizicheskoj Observatorii, 41

Borisenko A.N., Vitkovskij V.V., Zhelenkova O.P., et al., 1990, Bulletin of the Special Astrophysics Observatory, 32, p. 128

Derriere S., Gray N., McDowell J.C., et al., 2004, ASP Conf. Ser., 314, p. 315

Kononov V.K., Monosov M.L., Vitkovskij V.V., et al., 1990, Soobshcheniya Spetsial'noj Astrofizicheskoj Observatorii, 65, p. 32

Shergin V.S., 2014, Software for automatic calibration of coordinates of direct images, Tech. rep., SAO RAS

Vitkovskij V.V., Kononov V.K., Lipovetchkij V.A., et al., 1987, Development of the SAO AS USSR Archive, Tech. rep., SAO AS USSR

Vitkovskij V., Zhelenkova O., Kalinina N., et al., 2000, Baltic Astronomy, 9, p. 578

Vitkovskij V.V., Zhelenkova O.P., Ryadchenko V.P., et al., 1988, Soobshcheniya Spetsial'noj Astrofizicheskoj Observatorii, 59, p. 60

Warmels R.H., 1992, ASP Conf. Ser., 25, p. 115

Wells D.C., Greisen E.W., Harten R.H., 1981, Astronomy & Astrophysics Supplement, 44, p. 363

Zhelenkova O., Vitkovskij V., Chernenkov V., et. al., 2020, ASP Conf. Ser., 527, p. 69

Zhelenkova O.P., Vitkovskij V.V., Kalinina N.A., et al., 2003, Digital Libraries: Advanced Methods and Technologies, Electronic Collections, Proc. 5th All-Russian Conf., p. 359

Zhelenkova O.P., Vitkovskij V.V., Malkova G.A., et al., 2004, Digital Libraries: Advanced Methods and Technologies, Electronic Collections, Proc. 6th All-Russian Conf., p. 145

Zhelenkova O.P., Vitkovskij V.V., Plyaskina T.A., 2010, International Virtual Observatory—results of the first decade, Digital Libraries, Proc. Meeting, 13

Zhelenkova O.P., Vitkovskij V.V., Plyaskina T.A., et al., 2007, Open Internet access to the general archive of observations of the SAO RAS in the context of virtual observatory standards, Tech. rep., SAO RAS

Zhelenkova O.P., Vitkovskij V.V., Plyaskina T.A., et al., 2017, System synthesis and applied synergetics, Proc. All-Russian Conf., p. 143

Zhelenkova O.P., Vitkovskij V.V., Plyaskina T.A., et al., 2019, ASP Conf. Ser., 521, p. 128

Zhelenkova O.P., Vitkovskij V.V., Plyaskina T.A., et al., 2023, Acta Astrophysica Taurica, 4, p. 9