Construction and Application of the Xinjiang Astronomical Observatory Data Center Postprint

Authors: Zhang Hailong, Wang Jie, Wang Wanqiong, Nie Jun, Ye Xinchen, Zhu Yan, Tohtinur

Date: 2017-09-26T00:00:00+00:00

Abstract

Through years of observations, the Xinjiang Astronomical Observatory has accumulated a vast amount of valuable observational data. How to efficiently manage the rapidly growing massive observational data, achieve the fusion of full-band observational data, and archive and publish the massive data that will be generated by future larger-aperture telescopes are urgent issues that the Xinjiang Astronomical Observatory currently needs to address. The Xinjiang Astronomical Observatory Data Center provides astronomical scientific data services and infrastructure for astronomical research needs, and is operated and maintained by the Computer Technology Laboratory of the Xinjiang Astronomical Observatory. Constructed based on the latest virtual observatory-related standards, its fundamental services include: providing data storage, management, and publication services for astronomical observation projects; offering long-term preservation and access services for valuable astronomical scientific data and reprocessed data; and providing users with scientific data archiving and publication services along with related technical support. The primary data resources consist of: pulsar observation data from the Xinjiang Astronomical Observatory' s 25 m radio telescope; active galactic nuclei observation data; and molecular spectral line observation data. The data services provided encompass: PPMXL catalog cone search; online cross-matching of massive catalog data; unified content descriptor information query, among other services. This work primarily introduces the construction of the Xinjiang Astronomical Observatory Data Center and its services, how to use the published services for online data retrieval, and how to implement data operations using standard virtual observatory tools.

Full Text

Preamble

Astronomical Research & Technology, Vol. 14, No. 1, Jan. 2017
DOI: 53-1189 1672-7673

Construction and Application of the Xinjiang Astronomical Observatory Data Center

Authors:
Zhang Hailong[1], Wang Wanqiong[1], Nie Jun[1], Ye Xinchen[1], Zhu Yan[2], Tohtinur[1]

Affiliations:
[1] Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, China

Corresponding Author Email: zhanghailong@xao.ac.cn

Abstract

After years of observation, the Xinjiang Astronomical Observatory (XAO) has accumulated a massive volume of valuable observational data. The urgent challenges facing XAO include how to efficiently manage this rapidly growing data, how to achieve integration of whole-band observational data, and how to archive and publish the massive data that will be generated by future larger-aperture telescopes. The XAO Data Center provides scientific data services and infrastructure for astronomical research needs, operated and maintained by the Observatory's Computer Technology Division. Constructed based on the latest Virtual Observatory standards, its basic services include: (1) providing data storage, management, and publication services for astronomical observation programs; (2) offering long-term preservation and access services for valuable astronomical scientific data and secondary processed data; and (3) providing users with scientific data archiving and publication services along with technical support. The primary data resources include pulsar observation data, active galactic nucleus observation data, and molecular line observation data from XAO' s 25-meter radio telescope. Data services provided encompass PPMXL catalog cone search, massive catalog data online cross-validation, and Unified Content Descriptor (UCD) information query services. This paper primarily introduces the construction of the XAO Data Center and its services, and explains how to use the published services for online data retrieval and how to utilize standard Virtual Observatory tools for data operations.

Keywords: Data center; Virtual Observatory; Data release; Retrieval; PPMXL

1. Introduction

Countries worldwide have increased research and investment in advanced observational facilities. Several large astronomical telescopes have been constructed or initiated, including LAMOST (Large Sky Area Multi-Object Fiber Spectroscopy Telescope), which obtains tens of thousands of spectra per night; the 500-meter Aperture Spherical Radio Telescope (FAST); and the solar radio spectrometer that simultaneously achieves high spatial, temporal, and frequency resolution observations of solar dynamic processes. The Square Kilometer Array (SKA) and its precursor arrays such as ASKAP, LOFAR, and MeerKAT have already generated massive astronomical data volumes. The rapid development of computer network technology provides strong support for sharing massive astronomical data and Virtual Observatory technologies. The exponential growth of data poses severe challenges for database queries and cross-validation techniques. Traditional data access applications must now handle unprecedented pressures from massive astronomical data.

Full-text retrieval technology for searching keywords in document libraries has made significant progress. However, astronomical data retrieval faces special constraints. Standard files include two data formats: ASCII and FITS, both subject to two fundamental limitations: files are unsuitable for parallel processing, and metadata from merging different data types lacks precision. These limitations determine the unique characteristics of astronomical data retrieval.

Taking pulsar observations as an example, the Parkes 64-meter radio telescope in Australia has unique geographical advantages for long-term observations of the Galactic Center and southern Galactic latitude objects. As the only large radio telescope in the Southern Hemisphere, Parkes has discovered over half of all known pulsars. The Parkes pulsar database contains observation data from 1997 onward, including pulsar search data and pulsar timing data, with more than 100,000 data files publicly available for free. For timing mode observations, besides raw data, the database also includes data fully integrated in polarization and frequency. Downloading compressed files takes less time than downloading raw data. The Parkes data release website provides services based on PSRFITS, Timer, etc. To date, Parkes has conducted timing observations of 366 pulsars for 7 years, studying pulsar period glitches, proper motion, profile mode changes, and other phenomena. In 2006, a single-pulse observation mode was developed to study giant pulses from pulsars and radio emission from magnetars. These research projects involved multiple observers, and over ten years of observational data were stored dispersedly, often on personal and group computers. Currently, some data have passed their protection period, making it essential to build a data release and retrieval system to open pulsar data to researchers worldwide.

Many scientific achievements result from reprocessing historical data. The Parkes multibeam survey data have been reprocessed multiple times, and re-

analysis of pulsar survey data for the Large Magellanic Cloud discovered approximately 40 new pulsars and a strong, short-duration radio burst that may originate from outside the Milky Way. Many research results also require accumulation of data over years or even decades, such as studies of pulsar timing noise and period glitches. Therefore, building a complete data archiving and publication platform is essential to support data reprocessing and new discoveries, improve data utilization, and lay a solid foundation for adapting to international astronomical development trends and generating greater scientific value.

---

2. Data Center Construction

The XAO Data Center is accessible at two URLs that point to two mutually backed-up data storage servers: http://data.xao.ac.cn (located at XAO headquarters) and http://vo.xao.ac.cn (located at Nanshan Observation Station). The data center homepage is shown in [Figure 1: see original paper].

The configuration information for the two data servers is listed in , currently meeting XAO' s data archiving and publication needs. Each server features dual Intel Xeon E5-2692 v2 CPUs @ 2.20 GHz, 64 GB memory, 48 TB storage, Intel C600/X79 chipset, and 56 Gbps IB network cards.

2.1 Data Center Services

2.1.1 Astronomical Data Query Language (ADQL) ADQL is similar to SQL and is used to query astronomical catalogs or other tabular datasets. The service URL is http://data.xao.ac.cn/*system*/adql/query/form, which serves as the main query page. Through this page, users can query all data published by XAO.

Example 1:

```
select name,code from obscode.data where name like '%Nanshan%'
and code like '%1'
```

This queries the published observatory coordinate service to find stations whose names contain the string "Nanshan."

Example 2:

```
select * from ppmxl.main where 1 = CONTAINS(POINT('ICRS', raj2000,
dej2000), CIRCLE('ICRS', 10.6847929, 41.2690650, 0.1))
```

This retrieves all targets within a 0.1° radius cone around M31 (10.6847929, 41.2690650) from the PPMXL catalog.

2.1.2 Cone Search Cone Search is a data access protocol in the Virtual Observatory system for querying within a conical sky region of specified radius. It requires three input parameters: Right Ascension (RA, J2000), Declination (DEC, J2000), and search radius (SR), all in decimal degrees. Through an

optional parameter, users can specify the number of columns returned in tabular form. Cone Search aims for simplicity, enabling data publishers to implement it quickly as a powerful search tool that responds to HTTP GET requests.

The PPMXL catalog Cone Search service URL is http://data.xao.ac.cn/ppmxl/q/cone/form. Users input coordinates or object name, search radius, and click the Search button to view results.

2.1.3 Table Access Protocol (TAP) TAP defines a service protocol for accessing tabular data, including astronomical catalogs and general database tables. It provides access to database and table metadata as well as actual table data. TAP supports multiple query languages including ADQL and the developing Parameterised Query Language (PQL). TAP V1.0 provides special support for spatial extensions and indexed queries in ADQL, and supports both synchronous and asynchronous data query operations.

The TAP query service published by XAO is shown in [Figure 4: see original paper], enabling direct queries of data segments registered with the International Virtual Observatory Alliance (IVOA).

2.1.4 Simple Application Messaging Protocol (SAMP) SAMP enables interoperability and data exchange between astronomical software. It is impractical to build a single tool that meets all user needs, but protocols can constrain software behavior and interfaces. SAMP is an efficient protocol that leverages limited resources to enable individual tools to work better together. Its key feature is defining common file formats to complete data exchange between different applications. The core component is a messaging system that allows applications to share data and utilize each other' s functionality, supporting communication between desktop and web browser applications.

All data services published by XAO support SAMP, enabling query results to be directly passed to standard Virtual Observatory tools like TOPCAT for data mining and visualization operations.

2.2 Cross-Validation Service

Cross-validation is fundamental for multi-band data fusion and is critical for source type identification during target selection for large telescope observations. It involves matching sources between two different band catalogs: for each source in one catalog, the system searches for counterparts in another catalog within a certain radius. If the angular distance between two sources is smaller than a specific threshold, they are considered the same celestial object.

2.2.1 Principle For two points A and B on a sphere from catalogs A and B, with coordinates (,) and (,) respectively, the spherical angular distance can be calculated as follows. When the angular distance between two points is small, the cross-validation success criterion is:

$$d^2 = (\alpha_1 - \alpha_2)^2 \cos^2(\delta) + (\delta_1 - \delta_2)^2$$

When the distance between two points satisfies this formula (where r represents the error radius of the two catalogs), the points are considered successfully matched counterparts.

2.2.2 Implementation XAO' s online cross-validation service, shown in [Figure 6: see original paper], currently supports uploading table data files via local files or remote URLs. Users select a target catalog from all tables published by XAO, specify the RA (ucd: pos.eq.ra) and DEC (ucd: pos.eq.dec) columns, and input a search radius based on their instrument' s specific parameters. The service then returns cross-validation results.

Using the PPMXL catalog as a test target with 4,000 input conditions and a 0.001° radius, matching results can be obtained within seconds.

2.3 XAO Observation Data Services

The data center has essentially completed online storage of scientific data from the Nanshan 25-meter radio telescope, including pulsar, active galactic nucleus, molecular line, and intra-day variability (IDV) data, providing data release and access services to relevant researchers system-wide.

Taking pulsar data as an example, the pulsar data retrieval URL is http://data.xao.ac.cn/pul/pulsar/q/form. The system implements cone search and multi-constraint target retrieval based on metadata information such as observation target. Opening the query page and entering "J0332+5434" in the "Target Object" field with "*~2014*" in the "Data_" field and "HTML" as the output format allows preview of pulse profiles and related information. Users can click "Preview(pav-DFTp)" to view larger pulse profile images. The "Send via SAMP" button in the upper left corner can send retrieval results to VO tools like TOPCAT for visualization operations. Detailed query methods are described in reference [15].

---

3. Output Formats

The data center supports multiple output formats: HTML, FITS Table, CSV, JSON, VOTable, and tar archives. For large data volumes with many files, tar format is recommended because browser response times may cause timeouts. The "Limit to" field constrains the number of output rows, while "More output fields" allows selection of additional fields for richer data information.

---

4. Virtual Observatory Tool Queries

XAO' s pulsar data and catalog cone search services have been registered with IVOA. Since "xao"is the abbreviation for Xinjiang Astronomical Observatory, all services published by XAO can be retrieved through VO tools like TOPCAT by searching for services containing "xao" . Detailed query methods are described in reference [15].

---

5. Summary

The XAO Data Center has essentially established data archiving and publication functions, implementing archiving and release of historical data based on Virtual Observatory protocols. Currently, 1 GB of pulsar data has been archived. For pulsar data, multi-constraint target retrieval based on observation date, target, and other parameters has been implemented. For IDV and star formation data, cone search has been established. The massive data online cross-validation service, based on parallel computing technology, enables high-speed online cross-validation, providing both local data upload and remote URL options for cross-validation with published catalogs. Test results show millisecond-level response times for 4,000-record calculations.

The ADQL service allows users to query all published data in the data center. The UCD query service enables lookup of unified content descriptor information, and the observatory code query service provides coordinate information for observatories worldwide. Data mining and visualization operations for the XAO Data Center require further improvement.

Acknowledgments

We thank members of the Astronomy Science and Technology Cloud Project for their support in constructing the XAO Data Center. Data center testing and preprocessing were completed on the XAO Taurus high-performance computing system.

References

[1] Cui Xiangqun. LAMOST project and its progress. *Progress in Astronomy*, 123-128.
[2] Nan Rendong, Li Huixian. Progress of FAST in science, technique and instrument. *Science China*, 1063-1074.
[3] Bailes M, et al. Science with ASKAP. *Experimental Astronomy*, 151-273.
[4] Johnston S, Taylor R. The LOFAR beam former implementation and performance analysis. *Lecture Notes in Computer Science*, 328-339.

[5] Jonas J L. MeerKAT—the South African array with composite dishes and wide-band single pixel feeds. *Proceedings of the IEEE*, 140-151.
[6] Cui Chenzhou, Zhao Yongheng. Architecture of Chinese virtual observatory. *Publications of National Astronomical Observatories of China*, 1522-1530.
[7] Yu Hongyong, Zhao Tiejun, Zheng Dequan, et al. Full-text retrieval research and system implementation. *Beijing: Beijing Normal University.*
[8] Zhao Qing, Sun Jizhou, Yu Ce, et al. An efficient file-oriented retrieval method. *Journal of Computer Applications*, 2096-2060.
[9] Cui Chenzhou, Li Wen, Yu Ce, et al. Parallel massive data oriented astronomical cross-match. *Astronomical Research & Technology*, 116-123.
[10] Wells D C, Greisen E W, Harten R H. FITS—a flexible image transport system. *Astronomy & Astrophysics Supplement Series*, 363-370.
[11] Manchester R N, Lyne A G, D' Amico N, et al. The Parkes Southern pulsar survey—I. Observing and data analysis systems and initial results. *Monthly Notices of the Royal Astronomical Society*, 1235-1250.
[12] Hobbs G, Miller D. The Parkes observatory pulsar data archive. *Publications of the Astronomical Society of Australia*, 473-480.
[13] Manchester R N. The Parkes observatory pulsar data archive. *Astronomical Society of Australia*, 202-214.
[14] Zhang Hailong. Research on efficient astronomical cross-identification for massive data. *Tianjin: Tianjin University.*
[15] Demleitner M, Wang Na, et al. Data retrieval from Xinjiang Astronomical Observatory' s pulsar data archive. *Astronomical Research & Technology*, 140-151.

Note: Figure translations are in progress. See original paper for figures.