



## Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: Chinese Astronomical Data Center  
Website: <http://explore.china-vo.org/>  
Certification Date: 25 October 2018

This repository is owned by: Chinese National Astronomical Observatories, MOST/CAS



# Core Trustworthy Data Repository Requirements

---

## BACKGROUND INFORMATION

### Context

R0. Please provide context for your repository.

#### **Repository Type. Select all relevant types from:**

Domain or subject-based repository

Institutional repository

National repository system; including governmental

Research project repository

#### **Other (please describe)**

### Comments

The National Astronomical Observatories, Chinese Academy of Sciences (NAOC) is the host institute of the WDS facility, the Chinese Astronomical Data Center (CASDC). The Center for Information and Computing at NAOC is the host division of CASDC, which also hosts the core R&D team of the Chinese Virtual Observatory (China-VO).

CASDC maintains data from nationwide distributed telescopes, and mirrors of several important international astronomical datasets. These data are from not only NAOC but also other observatories in China, including Purple Mountain Observatory, Shanghai Astronomical Observatory, Yunan Astronomical Observatory and Xinjiang Astronomical Observatory. Each of these institutes has several observation sites.

#### **Brief Description of the Repository's Designated Community.**

At the end of August 2018, the CASDC web portal has 20,400 registered users, including about 2400 users from research or education institutes, and more than 18,000 amateur astronomers.

CASDC focuses on services for professional researchers and students, and also has several public outreach projects. Most amateurs do not necessarily fully understand astronomical data, but the population is huge and they are providing important contributions to citizen science projects.

#### **Level of Curation Performed. Select all relevant types from:**



- A. Content distributed as deposited
- B. Basic curation – e.g. brief checking; addition of basic metadata or documentation
- C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

### **Comments**

Raw data from NAOC telescopes and from several other observatories are archived in CAsDC with brief checking for quality control. Some mid-processed data—for example, LAMOST spectral data—are also archived. Product data from several institutes are published on the CAsDC platform. A number of very large catalogs are reformatted to CSV or imported to databases without changing the content.

### **Outsource Partners. If applicable, please list them.**

CAsDC has several long-term collaborators. We work with many projects in advanced research and funding calls:

Tianjin University— key technology research and development.

Central China Normal University—education and public outreach.

Kunming University of Science and Technology—key technology research and development.

Computer Network Information Center, CAS (CNIC)—key technology research and development.

Alibaba Cloud Computing Inc. (Alibaba Cloud)—key technology research and development; Cloud platform provider.

Wutai Tech. Inc.—system design and deployment.

### **Other Relevant Information.**

A CAsDC booklet is shared in [Annex A](#).



## ORGANIZATIONAL INFRASTRUCTURE

### I. Mission/Scope

Compliance Level: 4

**R1. The repository has an explicit mission to provide access to and preserve data in its domain.**

CAsDC is the scientific data service and infrastructure of (NAOC, which is a key service of China-VO. We are aiming to meet user requirements for astronomical research and education.

CAsDC provides the following services:

1. Data storage, management and release on behalf of astronomical projects.
2. Long-term preservation and access for historical archives.
3. Research and development of astronomical data management applications.
4. International exchange and sharing of scientific data.
5. Data-driven education and public outreach.
6. User support.

Ref.: <http://explore.china-vo.org/aboutus.html>

NAOC rules on data archiving and open access are shared in [Annex B](#).



## II. Licenses

Compliance Level: 3

**R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.**

Most data in CAsDC are completely open and without any limitations imposed on users; for example, BASS (Beijing-Arizona Sky Survey), SCUSS (South Galactic Cap U-band Sky Survey), and CSTAR (Chinese Small Telescope ARray).

National astronomical data sharing policies and project-specified data policies are in operation and published.

LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope) has released five versions (prior, dr1, dr2, dr3, dr4) of public data, and there are another two version (dr5, dr6) datasets still under embargo. Each version is released after an 18-month period of embargo. People using LAMOST data have to accept its Data Policy

(<http://dr.lamost.org/ucenter/doc/lssdp>) and Publication Policy (<http://dr.lamost.org/ucenter/doc/lsspp>).

AST3 (Antarctic Survey Telescopes) and several other datasets are also embargoed due to the requirements of the CCAA (Chinese Center for Antarctic Astronomy). Researchers must obtain permission from the CCAA to access the data. CAsDC provides the system that authorizes people to access the data, and the authorization process can also be operated by the data owner. The license applied to data is the decision of the data owner.

NAOC rules on data archiving and open access are shared in [Annex B](#).



### III. Continuity of access

Compliance Level: 4

**R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.**

CAsDC is hosted at NAOC with support from the Chinese Academy of Sciences, Ministry of Science and Technology, National Natural Science Foundation of China, and other government departments. NAOC requires that all data generated by NAOC instruments should be archived at CAsDC.

The core team members of CAsDC belong to the Center for Information and Computing of NAOC, and are all permanent staff. Annual operation funding from NAOC supports essential hardware requirements and Internet connection. Additional large expenditure comes from other funding sources.

CAsDC has dedicated servers and disk arrays. A full data backup is available at CNIC. Since 2016, CAsDC has been using public Cloud services,—for example, Alibaba Cloud—for data storage and public access. A fully strategic collaboration agreement was signed between NAOC and Alibaba Cloud in January 2017.

NAOC rules on data archiving and open access are shared in [Annex B](#).



## IV. Confidentiality/Ethics

Compliance Level: 4

**R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.**

There are no ethics concerns with astronomical data. However, instrument teams may keep data for 18 months or longer before publishing them to the international community (e.g., LAMOST), the exception being when restrictions from law or formal notices exist (e.g., AST3). CAsDC follows the requirements of data owners and provides different ways to make sure that each dataset is accessible only to authorized users.

LAMOST data policy is shared in [Annex C](#).



## V. Organizational infrastructure

Compliance Level: 4

**R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.**

NAOC is the host institute of the WDS facility, the CAsDC. The Center for Information and Computing at NAOC is the host division of CAsDC, where 15 full-time employees are working and also 3 graduate students.

There is annual funding from NAOC for the maintenance of fundamental instruments. In the coming three years or longer, Alibaba Cloud will provide adequate Cloud resources and funding to CAsDC. There have been two National Natural Science Foundation of China (NSFC) funding grants during 2016–2018. A new NSFC funding grant is also in place for 2018–2021 (including direct funding of 2.4 million Chinese Yuan).

The cover page of the new NSFC funding grant is shared in [Annex D](#).





## VI. Expert guidance

Compliance Level: 4

**R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).**

China-VO and the NAOC–Alibaba Cloud joint laboratory both have expert committees. CASDC shares the core R&D team of China-VO, and also its science committee, working groups, data service standards and protocols. The team represents CASDC within the work of the International Virtual Observatory Alliance (IVOA) and tightly follows its development of data sharing requirements.

We attend the Astronomical Data Analysis Software and Systems and IVOA conferences every year to communicate with experts in the astronomical software development field. Dr. Chenzhou Cui, the Principal Investigator of CASDC, now serves as Deputy Chair of IVOA (<http://ivoa.net/about/contacts.html>) and Chair of the IVOA Education Interest Group. Dr. Cui is also the newly elected Vice President of IAU Commission B2 (International Astronomical Union, Commission of Data and Documentation; [https://www.iau.org/science/scientific\\_bodies/commissions/B2/](https://www.iau.org/science/scientific_bodies/commissions/B2/)).

We also host at least two workshops each year to demonstrate how to use our data query system and to obtain feedback from the community.

A list of those in the China-VO Science and Technology Expert Committee is shared in [Annex E](#). Members come from astronomical observatories and universities. Most of them are academicians and professors. Although CASDC does not have an international committee, we have many international collaborations and dataset exchanges.



## DIGITAL OBJECT MANAGEMENT

### VII. Data integrity and authenticity

Compliance Level: 3

**R7. The repository guarantees the integrity and authenticity of the data.**

All data are backed up at a different physical data center. All datasets are verified by md5 checksum. Most data do not need version control as they are the production data from data owners. LAMOST spectral data have multiple versions corresponding to changes of pipeline code, and updated information is provided by the scientific data team of LAMOST (e.g., <http://dr2.lamost.org/doc/update20160603>). In CASDC, previous versions of datasets stay in database with date tags, and only that with the latest updates is published to users.



## VIII. Appraisal

Compliance Level: 3

**R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.**

CAsDC attempts to implement IVOA standards as much as possible. We check if files related to astronomical catalogs are all archived and in place. The quality of data is ensured by the data owner. CAsDC maintains the data and provides a publishing system. We collect all the required information and generate metadata to create data tables and build data indexes.

Our data query system has been designed to bind metadata with data. The query portal is generated automatically from metadata, obtaining data descriptions from the identifier specified in the metadata and so on. CAsDC's query system also provides several data formats, including FITS, CSV, and VOTable. FITS format is the most commonly used format in Astronomy.

A screen capture of the SCUSS data query form is shared in [Annex F](#).



## IX. Documented storage procedures

Compliance Level: 3

### R9. The repository applies documented processes and procedures in managing archival storage of the data.

CAsDC is a distributed platform with distributed backups. We have several data centers, and data are stored across them. NAOC has three CAsDC-dedicated disk arrays as well as several servers. Alibaba Cloud and CNIC Cloud both provide very good systems to keep data usable and safe. The storage procedures are documented in a local GIT repository. Standard processes are being established. For now, we can make sure that each dataset has more than two copies online and offline.

The research and development plans for CAsDC are shared in [Annex G](#).



## X. Preservation plan

**Compliance Level: 3**

**R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.**

CAsDC keeps all data for an unlimited period, except when the data owner asks for them to be deleted (which has never happened). Our data centers make sure that storage is well maintained and in good condition. With the help of Alibaba Cloud and CNIC Cloud, our data now have even greater insurance to be preserved for the long term.

The data owners determine whether their data are published or not. If it is still under embargo, CAsDC will not expose them freely through our portal, rather they can be accessible only to authorized users. CAsDC has the right to copy and transfer the data among data centers.

NAOC rules on data archiving and open access are shared in [Annex B](#).



## XI. Data quality

**Compliance Level: 3**

**R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.**

CAsDC has appropriate expertise to address technical data and metadata quality, and ensure that sufficient information is available for end users.

CAsDC staff all have professional skills in programming and data management. All staff members have a computer science and technology background, and each has more than 10 years programming experience. Moreover, all staff have at least a master's degree in computer science or astronomical techniques. This is also the core team that builds IVOA standardized services and scientific database projects.

Data quality control is operated by the data owner. When an owner believes that their data has reached production level, they ask CAsDC to publish it. Metadata and document requested by CAsDC are all provided by the data owner. We then convert the information into a format that the systems recognizes, and check if the information provided correctly corresponds to the dataset; for example, the table column format description.

Acknowledgements are appreciated, while users get benefits from CAsDC or China-VO (they are the same team). Ref: <http://astrocloud.china-vo.org/s/static/aboutUs.html> (the Acknowledgements section).



## XII. Workflows

**Compliance Level: 3**

**R12. Archiving takes place according to defined workflows from ingest to dissemination.**

The main work flow is from the data owner to the repository, then to the query system. The data owner generates the data; the repository assigns the storage space and pushes the catalog to the database; archived data are then integrated into the query system. Metadata and related information are also archived with the data.

The research and development plans for CAsDC are shared in [Annex G](#).



### XIII. Data discovery and identification

Compliance Level: 3

**R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.**

CAsDC query forms are generated from the metadata in the data catalogs. Users can input a coordinate and search radius to find which datasets contain data in the target area. Users can also directly use a dataset and apply more filters to obtain details about stars/galaxies or other astronomical targets.

The available datasets are all listed on the CAsDC website, and users can read the data descriptions there also. Some of the CAsDC data interfaces are also published in the IVOA registry, and users can find our datasets via a keyword search (e.g., LAMOST).

Usually, a paper describing a CAsDC data production will be submitted to a journal, which provides a persistent way for data citation.

As an example, the LAMOST publication policy is shared in [Annex H](#).





## XIV. Data reuse

Compliance Level: 3

**R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.**

Content-oriented metadata is needed by the query system and other related services in CAsDC. FITS is the format that is most commonly used by the astronomy community. Furthermore, CAsDC provides support for CSV and VOTable, which are also supported by new astronomy software packages. FITS and VOTable formats both have headers to self-describe the data. CSV is very flexible, and related information is added to the data.

A query form for the CSTAR dataset is shared in [Annex I](#), where a user can query data and find metadata at the same time.



## TECHNOLOGY

### XV. Technical infrastructure

Compliance Level: 4

**R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.**

CAsDC contains more than 500TB of data, has 1.5 PB storage capability, 700 TFlops computing power, and more than 100 software. More than five distributed sub-centers of CAsDC have been built to satisfy the requirements of people in different areas, and to reduce data transfer among sub-centers.

The CAsDC Cloud platform is adapted for the community, with many self-developed new features, and also a query system and storage system.

IVOA standards and protocols are widely adopted in CAsDC, including Simple Application Messaging Protocol, ConeSearch, Simple Image Access protocol, Simple Spectral Access Protocol, and so on.

The distribution of CAsDC nodes across China is shared in [Annex J](#).



## XVI. Security

**Compliance Level: 3**

**R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.**

CAsDC is hosted at NAOC, where necessary network security devices are equipped. Unnecessary ports of CAsDC servers are all disabled. CNIC also scans the whole CSTNet (network of the CAS) periodically, and hence potential weaknesses are found quickly and fixed.

The close collaboration between CAsDC and Alibaba Cloud has resulted in security being much improved; especially, for data on Alibaba Cloud. Software and services are continually scanning the system to reduce threats, and can even prevent flood attacks.

A CSTNet security scan report from 2016 is shared in [Annex K](#) as an example..



## APPLICANT FEEDBACK

### Comments/feedback

We apologize that a number of supporting documents are in Chinese. Only Chinese versions exist for many non-scientific documents.