



Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository Strasbourg Astronomical Data Center (CDS) complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.

The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on August 7, 2014.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version: 2014-2017 | July 19, 2013
Guidelines Information Booklet: [DSA-booklet_2014-2017.pdf](#)
All Guidelines Documentation: [Documentation](#)

Repository: Strasbourg Astronomical Data Center (CDS)
Seal Acquiry Date: Aug. 07, 2014

For the latest version of the awarded DSA for this repository please visit our website: <http://assessment.datasealofapproval.org/seals/>

Previously Acquired Seals: None

This repository is owned by: **Strasbourg Astronomical Data Center (CDS)**
Observatoire Astronomique de Strasbourg
11, rue de l'Universite
67000 Strasbourg
FRANCE

T 33 3 68 85 24 10
F 33 3 68 85 24 32
E francoise.genova@astro.unistra.fr
W <http://cds.unistra.fr/>

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

Strasbourg astronomical Data Center (CDS) is dedicated to the collection and worldwide distribution of astronomical data and related information.

The CDS hosts the [SIMBAD](#) astronomical database, the world reference database for the identification of astronomical objects; [VizieR](#), the catalogue service for the CDS reference collection of astronomical catalogues and tables published in academic journals; and the [Aladin](#) interactive software sky atlas for access, visualization and analysis of astronomical images, surveys, catalogues, databases and related data.

The CDS mission is to:

- collect useful information concerning astronomical objects that is available in computerized form;
- upgrade these data by critical evaluations and comparisons;
- distribute the results to the international astronomical community;
- conduct research, using these data.

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

URL: <http://cdsweb.u-strasbg.fr>

The CDS cooperates with the French Space Agency [CNES](#), the European Space Agency [ESA](#), the European Southern Observatory [ESO](#), the US National Aeronautics and Space Administration [NASA](#) (with a long term collaboration with the Astrophysics Data System [ADS](#) and the NASA Extragalactic Database [NED](#)), astronomical academic journals, and with other data and service providers around the world such as the National Observatory of China ([NAOC](#)), the Inter- University Centre of Astronomy and Astrophysics ([IUCAA](#) Pune, India), the National Observatory of Japan ([NAOJ](#)), the Institute of Astronomy of the Russian Academy of Sciences ([INASAN](#)) and the South African Astronomical Observatory ([SAAO](#)). CDS hosts mirrors of NASA ADS and of the [Astronomy and Astrophysics](#) international journal.

CDS is a member of the [World Data System](#) of the International Council for Science ICSU and has thus been certified following WDS criteria.

DSA label is sought only for VizieR and Aladin. The Simbad database is continuously updated with new information extracted from academic publications. It seems to us that it is not suitable for such a label.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Inputs for Aladin and VizieR services are tables issued from astronomical academic journals such as Astronomy&Astrophysics ([A&A](#)), Astronomical Journal ([AJ](#)), Astrophysical Journal ([ApJ](#)), Monthly Notices of the Royal Academy Society ([MNRAS](#)), catalogues and image surveys supplied by international agencies and centers such as NASA, ESA, ESO and [CADC](#) (Canadian Astronomy Data Centre), researchs teams and individual researchers.

The data arriving at the CDS are obtained from reliable sources, agencies and large projects and/or attached to a refereed publication wich reference is given. They are kept and redistributed in format and with attached metadata allowing them to be examined and scientifically reused. As explained later, they are compliant with disciplinary standards.

A standardized README file is attached to each catalogue. It contains the description of the catalogue content and information about its origin. For data linked to a publication, the article reference, abstract and date are given plus a link to the publication.

Example of data linked to a publication:

<http://cdsarc.u-strasbg.fr/viz-bin/Cat?J/ApJ/703/L72>

Example of data linked to a telescope:

<http://cdsarc.u-strasbg.fr/viz-bin/Cat?cat=B%2Fchandra&target=readme&menu=on>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. The data producer provides the data in formats recommended by the data repository.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Formats supported by the CDS are mainly:

- FITS format for tables and images. The FITS standard (http://fits.gsfc.nasa.gov/fits_standard.html) is a widely used disciplinary format for astronomical data. It is described in more details in item 7.
- ASCII format for tabular data and catalogues.

There is an interface allowing astronomers to submit their catalogue and its description (see item 3) for ingestion and checks.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

3. The data producer provides the data together with the metadata requested by the data repository.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Submission forms allow the data producers to add metadata which will then be validated by CDS staff (researchers, specialized librarians). The data and metadata follow standards adopted by CDS.

The data producer can directly deposit the data by the use of submission forms:

<http://cdsarc.u-strasbg.fr/cgi-bin/Submit>

Documentation: <http://cdsweb.u-strasbg.fr/vizier/submit.htx>

The CDS also extracts tables from the journals and builds the required metadata.

For catalogues supplied by agencies or teams which generate big data, metadata are built according to the documentation supplied by the data producer. They are generally the object of discussions between data producer and CDS managers.

Metadata and standards for tables/catalogues (CDS), in agreement with the journal *Astronomy&Astrophysics* and other journals: <http://cds.u-strasbg.fr/doc/catstd.htx>

The VOTable standards of the International Virtual Observatory Alliance [IVOA](http://www.ivoa.net/documents/VOTable/) (output format):
<http://www.ivoa.net/documents/VOTable/>

Bibliographical reference standard: <http://cdsweb.u-strasbg.fr/simbad/refcode/refcode-paper.html>

Recommendations to the authors on the journal sites:

http://www.aanda.org/doc_journal/instructions/aa_instructions.pdf

<http://aas.org/authors/manuscript-preparation-aj-apj-author-instructions>

http://www.oxfordjournals.org/our_journals/mnras/for_authors/

Process diagram for data quality checks:

<http://cds.u-strasbg.fr/vizier-org/OAISTranslation.html> (Description of ViZieR pipeline)

Tools developed at CDS transform received data in a dedicated format. These tools verify the coherence of the data: number of lines, number of columns, column type (integer, float, short, char). These informations are available in the REAMDE file.

The persons in charge of data validation are specialized librarians. In case of problems, they discuss the issue with the data provider and/or CDS astronomers.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

Comments:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The missions of the CDS are explained in the following document:

<http://cdsweb.u-strasbg.fr/about>

Strasbourg astronomical Data Center (CDS) is dedicated to the collection and worldwide distribution of astronomical data and related information. One of its mission is to:

"collect useful information concerning astronomical objects that is available in computerized form"

All the distributed data are archived.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Data distributed by the CDS are the object of agreements with data producers (disciplinary journals, data producers of the discipline such as NASA, ESA, etc.)

Observational data are public according to a timetable established by the data producer. In very rare cases there can be a "proprietary period" during which its usage is reserved to the team which is producing the data. Data in VizieR linked to a publication are public when the paper is published even if the article itself is not yet in open access.

For example the data policy of the international Journal "Astronomy & Astrophysics":

http://cds.aanda.org/index/php?option=com_content&view=article&id=136&Itemid=200

states that:

It is mandatory for A&A authors to publish the data that are presented and discussed in articles and needed to reproduce the results. Archiving the data also increases the value of the article, and thus its impact in the community. Publication of the data, usually at the CDS (see below), should occur immediately upon acceptance of the article referencing them. Some common examples of data that must be archived are the measurements of radial velocities leading to the detection of planetary or stellar companions to stars, the photometric data used in asteroseismologic studies, etc. By data, we mean here not only primary observational material, but also tools of general interest such as catalogs, theoretical tables of lasting values, etc.

Whenever the primary observational data (e.g., the spectrograms that were used for determining radial velocities or redshifts) are archived at a facility such as ESO or HST and therefore publicly available, there is no need for

authors to provide them to A&A; in this case, we'll archive only the reduced data (i.e., the radial velocities and the reduced photometric data in the examples given above). When primary data presented in articles are not publicly available through an institutional archive (e.g., the IRAM spectroscopic data), the calibrated data will be archived at the CDS.

By contract with A&A, the CDS stores the data that are published in A&A articles and graciously puts them at the disposal of the global community. The data are also linked to the general purpose data mining tools developed at the CDS and to the published articles through the ADS. The CDS requires the data tables to be in ascii format and each table is accompanied by a readme.txt file that describes the table's content. The readme file format defines a standard that is used by all major astronomy journals. Primary data can also be archived at the CDS as graphics files in FITS format. This is of particular interest for spectrograms. At this point, no other formats than ascii and FITS are supported by the CDS for A&A data. Also by contract with the Journal, CDS provides help to A&A authors in order to prepare the archival files.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. The data repository applies documented processes and procedures for managing data storage.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data are stored on RAID level 5 or 6 disks and backup of these data are made at regular intervals. These backups are made in a building distant from the datasever in a daily way. A low level supervision of the services (state of controllers, supplies, logical, physical and virtual disks, fans, temperature, UPS, etc.) as well as a supervision of the high level services are made by Nagios probes and warn in real time the engineers in charge in case of critical alert due to a system failure.

Electrical installations, UPS (Uninterruptible Power Supply), cooling systems, firewalls, computers, networks, etc. are redundant to insure a high level availability of the data repository.

The VizieR service has 9 mirror sites to mitigate any technical failure , and insure the best possible availability of service: ADAC (Astronomical Data Archives Center, Japan), CADC (Canadian Astronomy Data Centre), University of Cambridge Institute of Astronomy (UK), IUCAA (Inter-University Centre for Astronomy and Astrophysics, India), INASAN (Institute of Astronomy of the Russian Academy of Science), NAOC (National Astronomical Observatories, Chinese Academy of Science), JAC (Joint Astronomy Centre, Hawaii), CfA (Center for Astrophysics Harvard University, USA), SAAO (South African Astronomical Observatory, South Africa).

The ALADIN service has a mirror site at IAS (Institut d'Astrophysique Spatiale, Paris, France) for some data.

References for documented process:

<http://cds.u-strasbg.fr/vizier-org/OAISTranslation.html> (Description of ViZieR pipeline, Procedures in use)

<http://cds.u-strasbg.fr/vizier-org/replication.png>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. The data repository has a plan for long-term preservation of its digital assets.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data storage format are long-lasting formats: FITS metadata for images, and other disciplinary standards (ASCII, FITS, standardized metadata) for tabular data. The use of these formats guarantees the reconstruction of information systems over time independently of the used technologies, ie their conservation on the long term. ASCII files are independent of the used SGBD technology.

FITS (Flexible Image Transport System) is the standard data format used in astronomy to store, transport and archive data files. Its flexibility allows it to be used for a large variety of data types: tables, images, spectra, time series.

- The first version of FITS was released in 1981. Its evolution follows the "once FITS, always FITS" rule, meaning that developments of the format must not invalidate former existing FITS files.

- A FITS file is made of one or more Header + Data Units. Thus, metadata and data are kept together, the metadata being stored in ASCII as a set of keyword/value cards.

These two key aspects make FITS a very-well suited format for archiving and long-term preservation purposes.

More information about FITS can be found at http://fits.gsfc.nasa.gov/fits_overview.html

The data redundancy on external sites guarantees access toward all internal risks.

We use as far as possible recognized sustainable open source software and systems (PostgreSQL, Linux OS, etc.) which are a guarantee of sustainability.

We also insure a regular migration of the used technologies as proven by the fact that CDS started in 1972 and has maintained its data holding and databases since then, including of course several major migrations.

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

Migration plan since 1972:

1972 - 1979

Server : IBM 360/65 of Meudon Observatory, unique computer in French astronomy

Storage : removable IBM 2314 diskpacs, 29 Mb

2 disks at the beginning, 5 disks at the end

Backups : half inch magnetic tapes, 1600bpi and 6250bpi

1979 - 1981

Server : IBM Computer of the CNRS in Orsay

Storage : IBM disks 3330 or 3340 (?)

Backups : half inch magnetic tapes 6250bpi

1981 - 1984

Server : Univac 1108/1110 of the CNRS computer in Strasbourg

Storage : Univac disks 2x80 mega words of 36 bits

Backups : half inch magnetic tapes 6250bpi

1985 - 1990

Server : Univac 1110 of the Paris-Sud University (Orsay)

Storage : Univac disks.

Backups : half inch magnetic tapes 6250bpi

1990 - 1995

Server : DEC 5400 station at the Strasbourg Observatory

Storage : SCSI disks

Backups : exabyte cartridges (2.5 Gbytes at the beginning)

1995 - 2006

Servers : Several SUN stations (SPARC technology) at the Strasbourg Observatory

Storage : SCSI disks

Backups : DAT cartridges. Daily incremental backups, Weekly full backups

2007 - today

Servers : Intel and AMD CPU servers running Linux (Debian, Ubuntu, CentOS, Scientific Linux OS)

Storage : SCSI, SAS and FiberChannel disks in RAID 1, 5 and 6

Backups : Managed at the observatory level on a server in another building

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Archiving takes place according to explicit work flows across the data life cycle.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The "workflow" manages the data life cycle with retention of data said "obsolete" (mainly tabular data). A mechanism was set up which allows one to keep track of the history on the distributed data. The main ingestion and modification stages on the catalogues metadata are logged, signed and dated.

Services in CDS are living information systems, and metadata can evolve in VizierR.

Catalogues become obsolete when the data producer provides a new version. "Obsolete" data keep its ID and remains accessible, with a link to the current version.

Exemple of obsolete catalogue:

<http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=sdss>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The responsibility of the access and the availability of the data are managed through agreements with the data producers.

- The availability is insured by several service mirrors.
- The availability of services is improved by a mechanism (GLU) which allows to redirect the Web links towards another service mirror when they fail (checks every 10 minutes).
- The reconstruction of a service in case of major crash is insured by redundancy of databases, mirror copies and archiving.
- A replication of critical data is being implemented in a different building.

Mirror sites map:

<http://cds.u-strasbg.fr/vizier-org/VizieRRepartition.png>

<http://cds.u-strasbg.fr/vizier-org/replication.png>

OAIS documentation:

<http://cds.u-strasbg.fr/vizier-org/OAISTranslation.html> (VizieR responsibility in the archival)

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. The data repository enables the users to discover and use the data and refer to them in a persistent way.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The published catalogues are indexed by a name that is unique, standardized and reserved. This name is in agreement with the reference used in journals or for the other catalogues with a name as NN/DDDD (NN=roman numeral from I..X according to the subject of the catalogue, DDDD=sequential number).

The catalogue nomenclature is persistent. Articles have their own DOIs.

Link explaining the nomenclature:

<http://cds.u-strasbg.fr/vizier/doc/catstd-2.htx>

Example of link for a CDS table: <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/558/A18>

Data are distributed in compliance with the standards of the discipline and are made available through the Astronomical Virtual Observatory. OAI harvesting is effective: the IVOA Registry of Resources is OAI-PMH compliant.

IVOA standards: <http://www.ivoa.net>

IVOA Registry: wiki.ivoa.net/wiki/bin/view/IVOA/IvoaResReg

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

User friendly search tool for catalog selection.

Great description (byte-by-byte) of records.

11. The data repository ensures the integrity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

A control of integrity is done on the tabular data. This control consists in an audit on the Postgres database of the CDS based on triggers. It updates a logs table containing the transactions Delete/Insert/Update (date, user, table, IP address, software, data before the update and possibly the request).

The procedure follows the diagram:

<http://cds.u-strasbg.fr/vizier-org/replication.png>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. The data repository ensures the authenticity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The on-line publishing of tabular data and images is realized by the qualified librarians, approved and validated by astronomers who make sure of the data quality.

The procedures are described in the document:

<http://cds.u-strasbg.fr/vizier-org/OAISTranslation.html> (Astronomers part in VizierR archival and Description of VizierPipeline)

The data input is realized through a secure ftp service (vsftpd: very secure ftp daemon). The deposit of data is done by creation of a directory in which we put files. This directory is invisible and is known only by his(her) creator.

Finally, a program watches the upload by sending an e-mail when data are added and all transactions are logged (vsftpd.log)

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The technical infrastructure of CDS explicitly supports the task and function described in a standard like OAIS.

The technical infrastructure of Vizier is in compliance with archival standard.

The following document describes procedures "à la OAIS":

<http://cds.u-strasbg.fr/vizier-org/OAISTranslation.html> or

<http://cds.u-strasbg.fr/vizier-org/OAISTranslation.pdf>

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. The data consumer complies with access regulations set by the data repository.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data licenses of producers are preserved.

Here a link to the Vizier licence available for the consumers:

http://cds.u-strasbg.fr/vizier-org/licences_vizier.html

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data provider name and reference of publication (when applicable) are attached to the data to allow users to reference the origin of data, in agreement with the accepted code of conduct of scientific research.

A good description of the code of conduct of scientific research can be found e.g. in the ethics statement of the American Astronomical Society:

<http://aas.org/about/policies/aas-ethics-statement>

The following paragraphs are particularly relevant to CDS activities:

"Proper acknowledgement of the work of others should always be given, and complete referencing is an essential part of any astronomical research publication. Authors have an obligation to their colleagues and the scientific community to include a set of references that communicates the precedents, sources, and context of the reported work. Deliberate omission of a pertinent author or reference is unacceptable. Data provided by others must be cited appropriately, even if obtained from a public database.

All authors are responsible for providing prompt corrections or retractions if errors are found in published works with the first author bearing primary responsibility.

Plagiarism is the presentation of others' words, ideas or scientific results as if they were one's own. Citations to others' work must be clear, complete, and correct. Plagiarism is unethical behavior and is never acceptable.

These statements apply not only to scholarly journals but to all forms of scientific communication including but not limited to press releases, proposals, websites, popular books, and podcasts."

Reviewer Entry

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

Accept or send back to applicant for modification:

Accept

Comments:

16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data consumers are asked to quote the origin of data. The origin of data is available in the Readme file.

The CDS is not proprietary of the data and is not required to check eventual wrong usage, so no legal action will be introduced in case of misbehaviour/misuse of data.

Data is openly available and measures such as termination of access are not feasible.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Should rather be "not applicable"