



Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: Strasbourg Astronomical Data Centre
Website: <https://cds.unistra.fr>
Certification Date: 5 February 2019

This repository is owned by: Strasbourg Astronomical Observatory



Strasbourg Astronomical Data Centre

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of Repository

Strasbourg astronomical Data Centre (CDS – Centre de Données astronomiques de Strasbourg) is dedicated to collection and worldwide distribution of astronomical data and related information.

The CDS develops and hosts the SIMBAD astronomical database, the world reference database for the identification and bibliography of astronomical objects; VizieR, the catalogue service for the CDS reference collection of astronomical catalogues, tables and other data published in academic journals; and the Aladin interactive software sky atlas and virtual observatory portal for access, visualisation and analysis of astronomical images, surveys, catalogues, databases and related data. It also provides a top-level service allowing users to cross-match astronomical catalogues, including very large ones, to identify common objects, and a portal to access its services.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

CDS main user community is the world-wide astronomical community. Users are the service end-users, which are scientists, and also the community of data service providers such as agencies which run the ground- and space based telescopes and academic journals which use the services for their own needs. For instance, observatory archives use them to enable query by astronomical object name to their databases which have the position in the sky as query parameter, the CDS services providing the object position deduced from the name. Users are also CDS data providers, which include these agencies, the journals and any astronomer, who can have data she/he produced held by CDS, in particular when they provide data attached to publications in academic journals.

The services are also used more generally by people interested in astronomy, in particular the amateur astronomers, and in astronomy courses.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

The CDS performs data curation at data level, adding value to data by adding metadata and/or checking and updating the metadata provided by the depositors when available, and checking data/metadata consistency. User feedback is also taken into account to identify errors in the data, which are corrected, keeping track of the corrections.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

CDS is hosted by Strasbourg Astronomical Observatory (UMR 7550, a Joint Unit of the Centre National de la Recherche Scientifique (CNRS) and of Strasbourg University), which is located at Strasbourg University. The CDS uses services provided by the University and the Observatory. The statutes of the CDS defines those memberships.

Software development and hardware provision, maintenance and operations are managed by the CDS IT team with

support from the IT team of Strasbourg Astronomical Observatory. The network connection is provided by Strasbourg University OSIRIS network, which uses the Strasbourg Metropolitan Area Network for Research and Higher Education and the National Research and Education Network RENATER.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Other Relevant Information.

The CDS was created in 1972 as “Centre de Données Stellaires” (Stellar Data Centre). Its remit was extended in 1983 to include all types of astronomical objects, and it then was renamed “Centre de Données astronomique de Strasbourg”, keeping the acronym which was already well known.

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, the bibliographic reference database for astronomy, 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), and the South African Astronomical Observatory (SAAO). CDS hosts mirrors of NASA ADS (the reference astronomical bibliographic data base, maintained by NASA at Harvard University) and of the “Astronomy and Astrophysics” international journal. Mirror copies of the CDS services are hosted by partners: the VizieR service has 7 mirror sites respectively at ADAC (Astronomical Data Archive Center, Japan), CADC (Canadian Astronomy Data Centre), University of Cambridge Institute of Astronomy (UK), IUCAA, NAOC, CfA (Center for Astrophysics Harvard University, USA), SAAO (South African Astronomical Observatory, South Africa).

The CDS has been one of the main actors in the development of the disciplinary interoperability standards, and one of the key players of the International Virtual Observatory Alliance (IVOA), which defines the disciplinary interoperability standards, since its creation in 2002. It has been the cornerstone of the coordination of Virtual Observatory (VO) activities in France since the inception of the French VO VO-France in 2004, and has been leading the European coordination of the VO development, the European Virtual Observatory Euro-VO, since 2006. CDS has also been a partner, on behalf of CNRS, of the three first projects funded by the European Commission in support to the development of the Research Data Alliance (RDA). It leads the French National RDA Node set up in March 2018 by the 4th RDA Europe project in collaboration with the CNRS Scientific and Technical Information Direction (DIST).

The CDS services are major building blocks of the astronomical data landscape, with more than 900.000 queries/day in

2017.

The CDS has been one of the Research Infrastructures listed in the National Research Infrastructure Roadmap since its first publication in 2008. It has been a member of the World Data System of the International Council for Science ICSU, now International Science Council ISC, since 2012. It was certified as a trusted repository by the Data Seal of Approval in 2014.

VizieR is both a collection of datasets, which began at the creation of the CDS in 1972, and a database system browsing the data, which was released in 1996. The SIMBAD database is continuously updated with new information extracted from publications and large astronomical object catalogues. Aladin is both a collection of reference images of the sky and the sky atlas application, which accesses images from the CDS and remotely through the Virtual Observatory. CDS preservation mandate is for VizieR data holdings. This is why in the application the criteria are discussed with respect to VizieR, although the three services are maintained on the long term. On 1 September 2018, VizieR contains 17 603 datasets. The yearly growth is about 1200 datasets. The public document "VizieR Processes" is used as a reference for this application.

Relevant Links:

CDS: <http://cds.unistra.fr/>

Strasbourg Astronomical Observatory: <http://astro.unistra.fr/en/observatoire/>

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

RDA : <https://www.rd-alliance.org/>

France RDA National Node : <https://www.rd-alliance.org/groups/rda-france>

CDS in the National Research Infrastructure Roadmap:

- 2008 version: French version : http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/62/2/feuille_route_tgir_2008_527622.pdf, pp. VIII & 18, English summary: http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/62/7/roadmap_TGIR_2008_527627.pdf, pp. 8

- 2018 version : and http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/70/3/Brochure_Infrastructures_2018_948703.pdf, pp. 115 & 122 (English translation not yet available)

VizieR: <http://vizier.u-strasbg.fr/vizier/>

Reference document "VizieR Processes" - : <http://cds.u-strasbg.fr/vizier-org/> (also provided with this application)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The CDS mission was defined when it was created in 1972 and has remained the same since then:

- Collect useful information concerning astronomical objects that is available in computerized form;
- Upgrade these data by critical evaluation and comparisons;
- Distribute the results to the astronomical community;
- Conduct research, using these data.

The CDS is currently a joint endeavour of the CNRS – INSU (CNRS Institut National des Sciences de l'Univers - Universe Sciences Institute) and of Strasbourg University. It was created in 1972 as a joint endeavour of INAG (Institut National d'Astronomie et de Géophysique) and Louis Pasteur University, which were succeeded by CNRS-INSU and Strasbourg University respectively.

As explained in R0, the CDS mission is recognized at the highest possible level in France. It is officially labelled as one of the Research Infrastructures of the French National Research Infrastructure Roadmap since its first publication in 2008. The Roadmap is established and periodically updated by the ministry in charge of research, currently the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI - Ministry of Higher Education, Research and Innovation). It is also one of the Astronomy & Astrophysics Services Nationaux d'Observation (SNO – National Observation Services) of the CNRS-INSU in the Processing, Archival and Dissemination category (AA-ANO-5), in which it appears with its Infrastructure de Recherche status. This label is provided by a national commission established by the

CNRS-INSU.

Relevant links (accessed 13 August 2018)

CDS mission statement: <http://cds.unistra.fr/about>

CDS in the CNRS-INSU: <http://www.insu.cnrs.fr/node/1182>

CDS in the National Research Infrastructure Roadmap:

- 2008 version: French version : http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/62/2/feuille_route_tgir_2008_527622.pdf, pp. VIII & 18; English summary: http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/62/7/roadmap_TGIR_2008_527627.pdf, pp. 8

- 2018 version : http://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/70/3/Brochure_Infrastructures_2018_948703.pdf, pp. 115 & 122

CDS as a Service National d'Observation (see e.g. the map): <http://www.insu.cnrs.fr/node/1233>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. Licenses

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

All the data held by the CDS are publicly and freely available. The CDS does not hold nor provide access to sensitive data.

The CDS, through Vizier and its other services, provides public data for science. It gathers data produced by the community (journals, observatories, sky surveys...). It is only the owner of the metadata and other value it adds to the data, such as metadata or plots combining data from different sources. These added value are openly accessible for usage. The data licenses of producers are preserved:

- The data coming from academic journals are tables or other data associated to publications. It does not include plots nor other data built by the publishers.
- The data coming from research agencies or organisations can be subject to licences. The licences are exposed in the ReadMe file which describes the data and is provided with it. Example: see section Copyright in <http://cdsarc.u-strasbg.fr/viz-bin/Cat?l/271>

Vizier rules of usage are explained in: http://cds.unistra.fr/vizier-org/licences_vizier.html, which also explains how to find Copyright/licence information . We have set the goal of improving the accessibility of our licensing information.

The rules of usage refer to the code of conduct of scientific research in astronomy. 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 this accepted code of conduct. 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>, which is used as a reference by CDS. The following paragraphs are particularly relevant to CDS activities and are cited in the CDS rules of usage:

"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."

Compliance monitoring:

The data consumers are asked to quote the origin of data (see http://cds.unistra.fr/vizier-org/licences_vizier.html). The origin of data is available in the Readme file.

Data provided by CDS is openly available and measures such as identification of wrong usage or termination of access in

case of misbehaviour/misuse of data would not be feasible.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

For the future, it should be considered to make the license information easier to find for users. Currently they have to go actively looking for it which might lead to many users not being aware of the conditions of use.

III. Continuity of access

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

CDS has been preserving the original data it has collected since its creation in 1972, as well as the metadata necessary for data assessment and reuse. Access to data has been provided since then, with significant evolution in the access methods due to technological evolution. For instance, copies of catalogues were originally requested via surface mail, with a cost associated to the provision and shipping of listings, magnetic tapes or punch cards. Now the data is freely available on the web, and the dataset content can be browsed through the VizieR service or any service giving access to data through the Virtual Observatory which is able to give access to individual data items (row from a table, image from an image dataset, etc.).

The plan is continuous preservation, which requires that CDS continues to exist and that its expertise is maintained at the

highest level on the long term.

As explained in R0 and R1, CDS impact is fully recognized to the highest possible national level in France by its inscription in the National Research Infrastructure Roadmap by the ministry in charge of research since the first version of the Roadmap. It also has full support from the two organisations in charge, CNRS-INSU and Strasbourg University, which both provide funding and staff.

The organisational structure of the CDS, as described in R5, ensures stability of most of the staff and of its expertise on the medium/long term, since they are hired on permanent government positions provided by either the CNRS or Strasbourg University. Medium/long term sustainability is thus provided.

In the unlikely case in which CDS would lose support and disappear, the natural structure to host and conserve its data holdings would be the CINES (Centre Informatique National pour l'Enseignement Supérieur – National Computing Centre for Higher Education <https://www.cines.fr/en/>), which has high performance computing and long term digital preservation missions at the national level for higher education (<https://www.cines.fr/en/overview/>).

In addition, the file structure of VizieR data is very simple (data and data description in a hierarchical file system which is the basis of the current ftp access, IVOA OAI-PMH description of the datasets including Dublin Core parameters), data is reusable as discussed in R14, and it would be easy to continue to ensure access through a generic or institutional data repository, in particular the one currently set up by Strasbourg University. It is also likely that the data could be hosted by one of the major agencies active in astronomy.

It must be noted here that beginning now a discussion with agencies on a continuity plan and possible transfer of the data holdings would weaken CDS position at the international level by suggesting a fragility which does not exist.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. Confidentiality/Ethics

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

As explained in R2, CDS data are not subject to disclosure risk. The disciplinary code of conduct is a guideline for CDS, which in particular provides information allowing proper referencing of the origin of data and data citation.

In some cases, data ingested in VizieR are subject to a period with restricted access. This is the case for instance for data attached to a journal article which is not yet published, or for the preparation of data provided by a research organisation which is not yet released. In the first case, data is made available taking into account the article publication date. In the second one, VizieR can provide temporary restricted access with a non-public URL protected by userid/password.

Relevant links:

Code of conduct from the American Astronomical Society ethics statement (see also R2):

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

Reference document “VizieR Processes”: <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

- Section 5.5 “Procedures for data archive not yet published”

- Section 5.2 “Procedures in use for data from national or international data centres”

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

V. Organizational infrastructure

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

As explained, the CDS is a joint venture of the CNRS-INSU and Strasbourg University. CNRS is a major national research organisation which carries out research in all fields of knowledge through 10 institutes, including INSU. Strasbourg is the second most international university city in France (after Paris). Today, the University of Strasbourg counts 42000 students, offers initial and further education in a wide range of academic fields and is an international player in scientific research. Both CNRS and Strasbourg University provide the highest possible stability for research/higher education in France, which is in addition supported by CDS inscription in the National Research Infrastructure Roadmap by the ministry in charge of research.

From its creation in 1972, CDS governance has been assisted by an international Scientific Council of 12 members (6 French and 6 foreign members) composed of high level scientists, which includes representatives of the two intergovernmental organisations which play a major role in astronomy in Europe, the European Space Agency ESA and the European Southern Observatory ESO, and of NASA, the US National Aeronautics and Space Administration. The Scientific Council meets every year to evaluate CDS strategy and activities, and reports to CNRS-INSU and Strasbourg University. It is also in charge of proposing a name for CDS directorship. CDS Director is designated by CNRS-INSU in concertation with Strasbourg University, for a five-year mandate.

The overall annual budget of CDS is around 3 M€. It is mostly affected to the funding of permanent staff on government positions, including researchers, IT specialists and specialized librarians, about one-third of each category. In addition to the permanent staff, contractors are hired when required, in particular for R&D and for the development of new functionalities and projects. In August 2018, staff is currently 38 people, of which 30 are on permanent positions. In addition, around 10 software development internship (2-5 months) are open each year to involve students from all university levels in R&D activities and in the development of new service functionalities. The interns are closely tutored by

permanent staff to support the learning process and to make sure that the results can be reused afterwards.

The CDS works as an integrated team with constant interaction between astronomers, librarians and software engineers. The continuous dialogue between science needs and technological capacities has been an essential element of CDS sustainability, by allowing it to remain relevant to evolving user needs over more than 45 years which saw significant scientific progress and significant technological evolutions, including the web revolution. The organisation as an integrated team also allows continuous update of staff knowledge about scientific and technological aspects and their evolutions. Regular meetings are organised to exchange knowledge. For instance, a monthly meeting gathers the software engineers and interns of CDS and more generally of the Observatory, to discuss novel technologies and on-going R&D.

CDS staff are very present in the meetings relevant to the CDS activities, in particular in the yearly Astronomical Data Analysis Software and Systems (ADASS) conferences and in the Library and Information Services in Astronomy (LISA) conferences, which allow CDS IT specialists, librarians and researchers to present their work, to be aware of the most recent advances in their domains and to discuss with colleagues from over the world. The 2017 LISA conference, “Astronomy Librarianship in the era of Big Data and Open Science”, which was the 8th, was organised in Strasbourg by CDS and Strasbourg Observatory.

As explained in R0, CDS staff plays a key role in the IVOA. Two CDS staff are IVOA Executive Board members representing the European and French VO initiative respectively. The Executive Board Chair is currently (September 2018) the CDS director. CDS staff participate actively in the definition of the IVOA interoperability standards, and many are or have been Working Groups or Interest Group chairpersons. In August 2018, they have been editors or authors of 20 of the 44 IVOA standards.

Finally, CDS has been involved in the development of the Research Data Alliance since its definition phase. In September 2018, one of the co-chairs of the RDA Technical Advisory Board is a CDS member, who also co-leads the French RDA National Node.

Relevant links:

CNRS: <http://www.cnrs.fr/index.html>

Strasbourg University: <http://www.en.unistra.fr/index.php?id=21304>

CDS staff: <http://cds.unistra.fr/staff>

CDS Scientific Council membership: <http://cds.unistra.fr/staff#scientific-council>

ADASS site: <http://www.adass.org/>

LISA VIII conference site: <http://cds.unistra.fr/meetings/Lisa8/>

LISA VIII Proceedings : <https://www.epj-conferences.org/articles/epjconf/abs/2018/21/contents/contents.html>

IVOA site : www.ivoa.net

RDA site: <https://rd-alliance.org>

France RDA National Node : <https://www.rd-alliance.org/groups/rda-france>

Article published in the LISA VII Conference Proceedings which describes how librarians (“documentalists”) work with

astronomers and IT specialists at CDS:

Perret, E. et al.: Working Together at CDS: The Symbiosis Between Astronomers, Documentalists, and IT Specialists, ASP Conference Series, Vol. 492. San Francisco: Astronomical Society of the Pacific, 2015, p.13

<http://aspbooks.org/custom/publications/paper/492-0013.html>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

VI. Expert guidance

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

As explained in R5, the CDS has in-house experts in astronomy and IT. Astronomy has been and still is at the forefront of science data sharing, and CDS has been one of the early pioneers. This allowed the staff to develop a high level of expertise in data sharing. Staff members are often invited to act as experts at the national, European and international levels, in the astronomy domain or beyond. The first generation of CDS members has now retired, and continuity and further development have been fully ensured by new recruitments in the three staff profiles.

Astronomers' recruitment has been targeting active scientists covering a palette of different expertise in astronomy. All

CDS scientists are active in their research activities in parallel to their data centre duties. Some CDS scientists orient their research work towards science data sharing, but keep an up-to-date expertise of astronomy and its evolution. CDS astronomers are providing science requirements for the service development and maintenance, perform tests of the developments, are reference persons for the librarians for science questions, and work in close collaboration with the IT engineers and the librarians. Each service has attached astronomers, but CDS astronomers can provide input to all the services when required. Complementary scientific expertise can be sought from astronomers from the other team of Strasbourg Observatory when useful.

As explained in R5, the CDS Scientific Council provides high level feedback on CDS strategy and activities.

The active participation in ADASS and LISA conferences and in IVOA and RDA provides direct access to good practices developed outside of CDS, in astronomy and other disciplinary fields, and lively connection to the wider data sharing community.

Feedback is also sought from the community, from presentations in meetings such as those cited in R5, and through demonstration booths organised in science conferences, for instance at the yearly Winter meeting of the American Astronomical Society. CDS services are also included in many of the VO science tutorials. These tutorials are used in particular in yearly Schools organised at the European level. The Schools are targeted on early career astronomers, and include significant time dedicated to supporting the use of the services and tools by the participants for their own research projects. The tool and services developers and CDS science staff serve as tutors, and take note of the feedback and requirements.

In addition to these face-to-face interactions, each page of the CDS web has a “Contact” link (corresponding to the email address cds-question@unistra.fr) allowing users to ask a question or provide a comment. The feedback received from these links is managed by a subset of CDS staff, from which one person, on a weekly basis, takes the responsibility to dispatch each feedback to the relevant CDS person and to follow the response provided. It is also possible for users to post annotations on SIMBAD objects or VizieR datasets through the CDS Annotation service. For this they have to create a personal CDS account. The link to post an annotation is available on the object or dataset page. Annotations are regularly monitored and public feedback is added to the annotation when CDS has taken an action resulting from it.

Relevant links (accessed 13 August 2018):

List of CDS staff with their expertise domain: <http://cds.unistra.fr/staff>

Ref. document “VizieR Processes ” : <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

- Section 3.6.2 “Astronomers part in VizieR ingestion process”
- Section 3.6.6 “Services evolution and users feedback”
- Section 2.5 “The notion of “open” archive”

VO scientific tutorials: <http://www.euro-vo.org/?q=science/scientific-tutorials>

2017 VO European School: <https://www.asterics2020.eu/dokuwiki/doku.php?id=open:wp4:school3>

List of the last annotations posted by users: <http://cdsannotations.u-strasbg.fr/annotations/last?nb=100>

Article published in the LISA VII Conference Proceedings which describes how librarians (“documentalists”) work with astronomers and IT specialists at CDS (same reference as for R5):

Perret, E. et al.: Working Together at CDS: The Symbiosis Between Astronomers, Documentalists, and IT Specialists, ASP Conference Series, Vol. 492. San Francisco: Astronomical Society of the Pacific, 2015, p.13

<http://aspbooks.org/custom/publications/paper/492-0013.html>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

Each dataset (called “catalogue” in Vizier terminology) coming from journals is described in a ReadMe file:

<http://cds.u-strasbg.fr/doc/catstd-3.1.htx>

<http://vizier.u-strasbg.fr/vizier/submit.htx#ToC2>

The dataset content consists of one or several files containing tabular data or other types of “associated data” such as spectra, images, etc. For tabular data, the ReadMe file describes the tables which are stored in ASCII aligned columns. The ReadMe files contain the number of records and the byte-by-byte sections describing the format of each record. The ingestion process checks the consistency between the data and the ReadMe file.

Ref. document “VizieR Processes”

Section 2.1 “The archive”

Section 2.7 “Information provided by the producer (SIP)”

Section 2.8 “Archived information (AIP)” , section 3.4 “Data enhancement”

The ingestion process computes checksums (md5) for each new table coming from journals. The checksum is verified for the next updates (for example update in the metadata).

Checksums have been systematically computed since 2016 except for very large tables coming from telescopes and large surveys. The process will be generalized for the whole collection. Data transfers are done through rsync which computes checksums.

The history of ingestion actions (on data or metadata) are recorded into a private file which is included in the AIP. Each action is dated and signed by the performer (CDS staff).

Versioning is rare in VizieR: tables/catalogues depend of articles and do not have successive versions in general. A separate process is set up in case errors are detected. The CDS staff maintains relations with data producers (authors, editors, agencies, surveys). In case of doubt concerning the data, the CDS librarians in concertation with the CDS astronomers report the problem to authors/editors. The processes are described in more details in R8 (Quality control).

For catalogues subject to versioning -- e.g. large datasets coming from surveys -- VizieR creates a new identifier with a new ReadMe for the new version and puts the previous version to "Obsolete". The workflow manages the data life cycle with retention of data said "obsolete". 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. Example: SDSS catalogue: <http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=SDSS>

Reference document: “VizieR Processes” : <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

- Section 5.1.3 “Data transformation for input tables coming from journals”

- Section 5.2 “Procedures in use for data from national or international data centres”

As explained in R8, VizieR data comes from journal articles and recognized “authorities” such as observatories or large surveys. When the data does not come from a journal but is proposed directly by the author of a paper, CDS astronomers validate the data provenance, by checking the journal web site.

Authors can upload directly their data into Vizier submission tools

(<http://cdsarc.u-strasbg.fr/vizier.submit/publication-notes.html>) but this is just a first step. All those data are ingested into Vizier via the usual complete process and, so, are verified by documentalists and validated by astronomers who both have access to the complete articles.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

VIII. Appraisal

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

- Collection development policy

Vizier collection development policy is aimed at gathering high quality data, which is essential to maintain community trust in the service. To fulfil this aim it was decided that the policy is to accept data attached to a publication in a refereed journal or data provided by a ‘trusted’ project: an organisation running a space or ground-based telescope or one of the so-called “large surveys”. These surveys gather large international teams who have been granted significant allocations of telescope time at national or international observatories, and sometimes on a dedicated telescope or space mission. Such survey programs are extensively reviewed by funding agencies and/or observatories before they are accepted. In all

cases CDS establishes collaboration with the relevant journals, observatories or surveys.

Specific ingestion pipelines were established with three of the main astronomical journals. The scientific editor of “Astronomy & Astrophysics” decides which data should be included in VizieR and requests the authors to provide the data and draft metadata. The American Astrophysical Society editorial office, which manages the “Astrophysical Journal” and its Supplement Series and the “Astronomical Journal”, provides CDS with the data and draft metadata. “Monthly Notices of the Royal Astronomical Society” encourages the authors to submit their data to CDS (see last paragraph of Section 2.7 in https://academic.oup.com/mnras/pages/General_Instructions#2.7%20Catalogues%20and%20online-only%20material). More generally, CDS astronomers can decide that a dataset attached to a refereed publication would be a valuable community asset in VizieR and initiate ingestion, and authors can submit their data attached to an article to CDS. In the case of very large catalogues (typically larger than 107 rows) coming from observatories or large surveys, a CDS engineer arranges for transfer of the relevant data from the provider, and then formats the data in the CDS binary format and builds the metadata (the ReadMe file) with the help of the documentalists.

- Quality control

The ingestion pipeline and full data life cycle at CDS is described in “VizieR Processes”. The overall process concerns the creation and checking of metadata, and checking the consistency of the data and metadata. Draft versions of the metadata may be submitted by the data providers, but these are always checked by CDS librarians and validated by CDS astronomers before ingestion in the database and publication on the web site. A dialog is established with the data provider when useful. The metadata process is explained in more details below. Metadata are sufficient to ensure data reuse and long term preservation, with ASCII and FITS (http://fits.gsfc.nasa.gov/fits_standard.html), the basic astronomy data format, as basic components. The preferred formats allows data to be reused.

Metadata are checked and updated if necessary when they have been provided with the data, or they are prepared by CDS staff.

- For tabular data, the metadata provides a full standardized description of the table columns which is contained in the ReadMe file. The standard was defined by CDS and accepted by the community, including the journals. One of its aims is to link the physical structure of the file to its scientific content. It is used to ingest the files in VizieR database, and enables checks of the file content.

Standard description: <http://cds.u-strasbg.fr/doc/catstd-3.1.htx>

ReadMe example: <ftp://cdsarc.u-strasbg.fr/pub/cats/J/A%2BA/600/A129/ReadMe>

- Other types of data (image, spectra), the “associated data”, are provided in the disciplinary standard FITS, which includes metadata allowing data reuse, and the datasets are described with the relevant IVOA standard ObsCore (<http://www.ivoa.net/documents/ObsCore/>).

Quality control checks are also provided to users. The tools include format validation and consistency checks between metadata and data.

- <http://cdsarc.u-strasbg.fr/vizier.submit/publication-notes.html>

- <http://cdsarc.u-strasbg.fr/vizier.submit/help.html>

- <http://cdsarc.u-strasbg.fr/vizier.submit/>

When the data checks identify suspicious data (values inconsistent with the data type, object names not matching the position, etc.), CDS reports the problem to the author or to the journal editor. These reports can lead to an erratum published in the journal. With the author's agreement, VizieR tables are corrected. The ReadMe file keeps track of the correction and the original files are kept. If there is no answer from the author, CDS astronomers decide on what is ingested, ingestion as it is with a note, non-ingestion or correction of the values (with comments).

Correspondence with the depositors is archived (non-public) to capture full provenance of the records and changes.

- Non-conformant formats and metadata

The list of accepted formats is described in the documentation: it includes TSV, FITS (<https://fits.gsfc.nasa.gov/>), ascii aligned columns for tables, FITS for images or spectra.

For unexpected data format (for example binary files, jpeg, etc.), the CDS will choose to provide or not the data and to take in charge or not the preservation and migration, and will inform the data provider of the decision if it is negative. For tabular data, CDS encourages the usage of the recommended formats but is able to and does make table format migration when required.

Authors are encouraged to provide metadata (ReadMe for tabular data, ObsCore for other data) but they are always checked by CDS staff, which adds or completes them if they are absent or incomplete.

Reference Document "VizieR Processes" : <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

Section 3.6.2 "Astronomers part in the VizieR ingestion process"

Section 3.6.3 "Documentalist part in the VizieR ingestion process"

Section 5.1 "Procedures in use for data coming from journals", 5.1.3 "Data transformation for input tables coming from journals"

Section 5.2 "Procedures in use for data coming from national or international data centres"

Section 3 "Description of VizieR ingestion pipeline"

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

IX. Documented storage procedures

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Storage procedures are described in Section 3.3 “Archives storage” of the “VizieR Process ” document. Redundancy of data and metadata is provided locally and on remote mirrors. The storage system is distributed (two different buildings at the Observatory, 7 mirror sites) to allow recovery in case of a major crash. The metadata allow to rebuild VizieR Information System since they are the basis for data ingestion in the database. The VizieR service hosted at CDS is enabled to switch to a mirror if the local database is not available.

- Storage architecture

The data are stored on RAID level 5 or 6 disks. Daily backups of these data are made to a backup system in a building distant from the data server. 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 power line, UPS (Uninterruptible Power Supply), cooling systems, firewalls, computers, networks, etc. are redundant to ensure a high level availability of the data repository.

- Monitoring and failure management

- The CDS supervision (internal usage) detects the CDS software failures: http access, database monitoring.
- The VizieR supervisor monitors the mirroring process and advanced software failures (nagios)
- The CDS nagios server detects hardware failures

The Strasbourg Observatory System Team intervenes in case of hardware failure. All hardware are under contracts with

the hardware supplier to a d+1 intervention.

An emergency manual is available in paper and electronic versions. The document is used for intervention in case of service crash or disturbances. It can be used by all CDS service managers or staff from the Observatory IT team to intervene on any service.

- Mirrors

As explained in R0, the VizieR service has 7 mirror sites to mitigate any technical failure, and to ensure the best possible availability of the 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), NAOC (National Astronomical Observatories, Chinese Academy of Science), CfA (Center for Astrophysics Harvard University, USA), SAAO (South African Astronomical Observatory, South Africa).

The mirror synchronization is based on rsync for data files and web applications. The mirroring database process is incremental with a table granularity. The database mirroring process synchronizes catalogues including metadata and indexation. Each mirror has at least a copy of the database. The Indian mirror (IUCAA) contains a complete mirror of all large tables (which are not stored in the database). The Japanese mirror (ADAC/NAOJ) contains the whole public archive.

References:

“VizieR Process ” document: <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The CDS has been preserving its data holdings since its creation in 1972. Elements of the preservation plan are included in the "VizieR Processes" document. The essential aspect is the fact that the data formats and metadata are adapted for long term preservation and enable data reuse.

The data storage formats are long-lasting formats: FITS metadata for images, and other disciplinary standards (ASCII, FITS, standardized metadata) for tabular data. Very large tables are a specific case discussed below. The use of these formats guarantees the reconstruction of information systems over time independently of the used technologies, i.e. 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. It is now maintained under responsibility of the International Astronomical Union. 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

Very large tables are stored in a binary format specific to CDS. For these tables CDS is in charge of dissemination and added-value provision, not to preserve the original data. The data has nevertheless be preserved in CDS format and kept usable over time. The format is fully documented in the code maintained by CDS.

CDS has been constantly monitoring its hardware performances to initiate updates when needed, preserving its capacity to respond to huge community usage. Along the CDS life time, a number of migrations have been performed, which are summarized here for information. This long history without usability loss demonstrates the format/metadata robustness to migration.

1972 – 1979

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

Storage : removable IBM 2314 diskpacks, 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 centre 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

2014: Data replication on another building of the Observatory

References:

“VizieR Processes” document: <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

- Section 3.7 “Preservation planning”

- Section 4 “Responsibilities for managing VizieR”

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XI. Data quality

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The information about how CDS manages data quality is available in R8. The initial quality assessment is provided by the journals, observatories and surveys from which the data originates. Quality checks are performed by CDS staff in a similar way for all the datasets deposited at CDS. Data/metadata consistency is checked. Metadata prepared or checked and updated by a CDS documentalist are validated by an astronomer before publication. In addition to the metadata provided by CDS, the README contains links to associated documentation. Most datasets in VizieR are associated to a publication which provides valuable additional information. Link to the original publication is provided, together with links to additional information, for instance to the project which produced the data if relevant (example in <http://cdsarc.u-strasbg.fr/viz-bin/Cat?J/ApJ/830/127>).

As explained in R6, it is possible for users to post annotations on SIMBAD objects or VizieR datasets through the CDS Annotation service. For this they have to create a personal CDS account. The link to post an annotation is available on the object or dataset page. Annotations are regularly monitored by CDS staff and feedback is added to the annotation when CDS has taken an action resulting from it.

References:

Reference document “VizieR Processes” : <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

- Metadata contents: Sections 2.7 “Information provided by the producer (SIP)”, 2.8 “Archived information (AIP)” and 2.9 “Distributed information (DIP)”

- Pipeline and metadata assignment processes: Section 3.1 “VizieR pipeline”, 3.2 “Data reception”, 3.3 “Archives storage”
- Section 3.6.2 “Astronomers part in the VizieR ingestion process”
- Section 3.6.3 “Documentalists part in the VizieR ingestion process”
- Section 5.1.3 “Data transformations of input tables coming from journals”

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XII. Workflows

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The organisation of VizieR archive and the associated workflows are described in the “VizieR Processes ” document. The document describes how the different pipelines interact with the CDS staff and tools. It is inspired from the OAIS reference model (without asserting that the VizieR archive is a fully compliant OAIS system). The environment of the VizieR archive in the OAIS workflow (Producer-OAIS archive/Management-Consumer) is characterized in Section 2. VizieR pipeline is described in Section 3 and a schema summarizing the full VizieR workflow is provided p.14 . The building blocks correspond to the Entry, Archival storage, Data Management, Administration, Access and Preservation planning OAIS entities. Different aspects have been described in the answers to other criteria, in particular R8 for ingest, R9 for storage

and R10 for data quality.

Constant attention is paid to the workflow efficiency. Detected problems are discussed in the weekly meetings of the VizieR team and in the CDS meetings which include the whole CDS staff. Contact is kept with the journals and the large data producers to identify and discuss problems, and specific meetings can be organised when required (for instance with journal editors).

Reference document: "VizieR Processes": <http://cds.u-strasbg.fr/vizier-org/> (provided with this application)

- Pipeline: Section 2 "OAIS environment applied to VizieR archives"
- Section 3 "Description of the VizieR pipeline ingestion processes"
- Section 5 "Procedures in use"
- Data management: Section 3.6 "Information system administration"

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIII. Data discovery and identification

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Enabling data discovery is one of the core CDS functions. The Vizier service (<http://vizier.u-strasbg.fr/viz-bin/VizieR>) enables users to find the data they need in the database. Datasets are discoverable by interfaces that are adapted to astronomical usage (for example with categories of wavelength bands, missions, astronomical topics) as well as through open keywords, which can for instance be author names or words from the title or description, or by sky position. All datasets containing data about a given quantity (for instance, “redshift” or “star formation rate”, which are relevant to galaxies) can also be retrieved, thanks to the systematic tagging of all the table columns with their Unified Content Descriptor (UCD – an IVOA standard describing quantities used in astronomy) (for example the ‘Search for catalogues by UCDS’ in the Vizier page).

VizieR is also fully available through the Virtual Observatory. All the datasets have an individual entry harvested by the IVOA Registry of Resources. This registry is based on the OAI-PMH protocol, the metadata contain the full Dublin Core parameters plus disciplinary extensions defined in the IVOA standard “Resource metadata for the Virtual Observatory”. All VO-enabled tools can thus discover the datasets.

One can note here that the IVOA Registry of resources is included in the generic registry developed by EUDAT, B2FIND, in which IVOA appears as one of the identified “communities” (<http://b2find.eudat.eu/group>). This is an example of harvesting made possible by the use of the OAI-PMH protocol.

Access to data is provided through the VO Data Access Layer protocols for tables, images and spectra. For instance, thanks to the tool and service interoperability enabled by the VO, the widely used TOPCAT tool (<http://www.star.bris.ac.uk/~mbt/topcat/>) can straightforwardly be used as a Vizier interface instead of the CDS web interface.

How to acknowledge usage of Vizier is explained in Vizier Rules of Usage.

“The usual rule is to cite the authors, the producer and the publisher of the catalogue; details can be found in the associated ReadMe file. A Vizier catalogue can be directly referred to as “VizieR Catalogue” followed by the Vizier unique identifying code, for example “VizieR Catalogue J/A+AS/124/75” or “VizieR Catalogue II/246”).”

All Vizier datasets have a unique identifier, which is 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 between I and X according to the subject of the catalogue, DDDD=sequential number).

Link explaining the nomenclature: <http://cds.u-strasbg.fr/vizier/doc/catstd-2.htm>

Example of identifier for a CDS catalogue published in a journal: <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/558/A18>

CDS also developed a bibliographic ID in collaboration with the ADS, based on the unique identifier of the initial publication or specific to the catalogue for surveys or data provided by an observatory. All Vizier datasets are referenced in the ADS, which is the usual source for reference citation used by the astronomical community.

CDS is in the final stage of preparing the attribution of DOIs to VizieR datasets. This DOI metadata will cite the initial article DOI for data attached to publications. At present there are several ways of citing VizieR catalogues in publications using VizieR unique identifier but when DOIs are implemented (around the end of 2018) they will be the recommended way for citation.

Examples of citations found in the literature:

In the bibliography section of 2017MNRAS.467.2517S:

Cutri R. M. et al, 2003, VizieR Online Data Catalog: II/246. CDS, Strasbourg, France

In 2017Ap.....60...70G, in the text:

"We used the JHKs photometry from the 2MASS catalog, the IRAS photometric data, which were used to select the original BIS (VizieR catalogue III/237A) sample, the observations of the AKARI satellite at 9 and 18 μm [16] and of the WISE satellite at 3.4, 4.6, 12, and 22 μm (VizieR Catalogue II/328)."

Relevant IVOA standards:

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

UCD – An IVOA standard for Unified Content Descriptors V1.1 <http://www.ivoa.net/Documents/latest/UCD.html>

UCD1+ Controlled Vocabulary V1.3 <http://www.ivoa.net/documents/UCD1+/20180527/index.html>

IVOA Registry Interfaces V1.1 <http://www.ivoa.net/documents/RegistryInterface/20180723/>

Resource metadata for the Virtual Observatory V1.12 <http://www.ivoa.net/documents/RM/20070302/index.html>

TAP - Table Access Protocol V1.0 <http://www.ivoa.net/documents/TAP/20100327/>

SIA – Simple Image Access Protocol V2.0 <http://www.ivoa.net/documents/SIA/20151223/>

SSA – Simple Spectral Access Protocol V1.1 <http://www.ivoa.net/documents/SSA/20120210/index.html>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIV. Data reuse

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The metadata attached to each dataset (ReadMe or ObsCore descriptions, see R8) are fully standardized and are designed to enable data reuse. For instance, for tables the description includes information about data types, formats, units, and explanation about the data which enable basic reuse, and provides information about the origin of the data. The FITS format includes information about the observation in addition to the data itself, and is fully reusable through many astronomy tools.

VizieR outputs are available in different formats used by the community, such as FITS (which is a format recognized by the Virtual Observatory), CSV (comma-separated values), and VOTable (<http://www.ivoa.net/documents/VOTable/20130920/>), which is a tabular format defined by the Virtual Observatory, which allows the data to be readily useable with VO-enabled tools. Large datasets are available in the VO HIPS format (Hierarchical Progressive Survey - <http://www.ivoa.net/documents/HiPS/20170519/index.html>), which allows fast hierarchical access to data with increasing positional resolution.

CDS ensures format migration as well as technological migrations, for instance when VO standards are updated. VizieR is both the dataset collection, which begun with the CDS in 1972, and a database system browsing the data, which was released in 1996. The timeline of VizieR-related migrations can be summarized as follows:

1996: VizieR initial release - C-language + Sybase data base management system. Large datasets have a specific ad-hoc management system

2002: Output to the Virtual Observatory from VizieR (VOTable format available as output)

2008: VizieR : PostgreSQL database management system

2012: Large dataset format version 2 (processed by Java process)

2014: VizieR table are queryable with ADQL/SQL (VO standard <http://www.ivoa.net/documents/TAP/>); large datasets are also stored into a transactional database

2015: Large dataset are also stored in HIPS

2016: Associated data are indexed in VizieR with the ObsCore data model (<http://www.ivoa.net/documents/ObsCore/>)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

TECHNOLOGY

XV. Technical infrastructure

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Systems and storage:

Our services rely on redundant infrastructures to ensure the highest availability and performances.

The CDS technical infrastructure is based on standard hardware (~40 DELL servers with Intel processors) and open source Linux operating systems (Debian, Ubuntu, CentOS, Scientific Linux OS).

Since 2018, we started to migrate some servers on a VMWare virtualized infrastructure for redundancy purposes and resource optimization.

Servers are dual attached to redundant ISCSI and FiberChannel RAID storage arrays.

Servers and storage arrays are hosted in two different rooms with redundant cooling and secure electrical supplies.

Software:

We tend to use mainly open source software. The Vizier service is currently implemented in the PostgreSQL database system (<https://www.postgresql.org/>), which is widely used in astronomy. Tools and libraries in Vizier are under free licenses. Software code development is managed by software version control.

Access to the services through the Virtual Observatory interoperability framework is implemented and updated to take advantage of new IVOA recommendations when relevant.

Management and documentation:

CDS is constantly monitoring infrastructure usage and implement upgrades when required, as shown in R10 and R14. On top of that, we use GLPI (<https://glpi-project.org/>) as IT management solution and software inventory. It collects information on the whole CDS infrastructure (hardware and software). We use a wiki for documentation and the critical procedures are printed and stored into an emergency binder.

Network connectivity:

CDS network connectivity is composed of a redundant dual attached high bandwidth (2 x 10 Gb/s) connection to the University Osiris Network (used for research and higher education) which relies on RENATER (the French NREN). The available bandwidth matches our current and further needs, it is monitored by our supervision solution (based on Nagios). Osiris and RENATER teams are really reactive and efficient.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XVI. Security

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The CDS commits to ensure security by implementing redundancies at every IT level. The CDS Information System (including VizieR archives) follows the current safety regulations of the CNRS and the University.

Organization and incident management:

The CDS is hosted by Strasbourg Observatory which is an Unité Mixte de Recherche of the Centre National de la Recherche Scientifique (CNRS) and of the Université de Strasbourg. CDS services security is coordinated by our local security officer who interacts directly with the Osiris CERT (Computer Emergency Response Team) driven by Strasbourg University CISO (Chief Information Security Officer) and CNRS Regional Security Manager. The Osiris CERT is working directly with the RENATER CERT, so it provides us with two different levels of incident detection and very high level engineers to support us if needed.

Incident detection and preventive actions:

CDS network is protected by two redundant high performance firewalls (2 x 10 Gb/s). A very strict filtering policy is implemented and we monitor connections to ensure that traffic is licit. CDS public servers are isolated in a specific high security network area. The firewalls warn us about security issue on our operating systems and software as they are able to analyse application data. We are thus able to define and implement software upgrades to correct the problems. Our security officer also uses Nessus (<https://www.tenable.com/products/nessus/nessus-professional>) to scan CDS servers to detect security vulnerabilities.

Secured hosting environment:

CDS infrastructure is hosted and replicated in two different Observatory buildings with secured facilities disposals:

- Rooms are secured with badge access control, alarms are directly connected to a supervisory agency. So only authorized people can access.
- Secured electrical supplies with UPS and redundant cooling system (with maintenance contract).

Technical failure prevention:

All CDS assets are covered by maintenance contract, mainly with a next day replacement for hardware failures. Failures are detected by the monitoring system. By design, CDS services rely on redundant infrastructures, so we can mitigate issues and reduce downtime.

Disaster recovery plan:

Our IT team is working on the definition of formal disaster recovery plan to identify threats and risks and define a way to prevent them. The plan will be available for mid-2019. We already have a strong infrastructure and good documentation, so the plan will mainly be a formalization of what is already implemented.

Redundancy design:

As we are able to host our servers and storage arrays in two different buildings, we can provide a very high level of availability for CDS critical services as data and computing power is redundant.

Data and software mirroring:

VizieR is also mirrored in different locations around the world to ensure their availability and preservation, as explained in R9.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

This CTS application follows our DSA Certification (2014-2017). The new procedures have motivated us to improve the description of our services. We also plan to maintain our WDS membership.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept