



IDOC/IDOC-DATA

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:

OK

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, National repository system; including governmental

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

OK

Brief Description of Repository

IDOC-DATA is a department of IDOC

IDOC (Integrated Data & Operation Center) has existed since 2003 as a satellite operations center and data center for the Institute of Space Astrophysics (IAS) in Orsay, France. Since then, it has operated within the OSUPS (Observatoire des Sciences de l'Univers de l'Université Paris-Saclay - first french university in shanghai ranking), which includes three institutes: IAS, AIM (Astrophysique, Interprétation, Modélisation - IRFU, CEA) and GEOPS (Geosciences Paris-Saclay) .

IDOC participates in the space missions of OSUPS and its partners, from mission design to long-term scientific data archiving. For each phase of the missions, IDOC offers three kinds of services in the scientific themes of OSUPS and therefore IDOC's activities are divided into three departments:

IDOC-INSTR: instrument design and testing,

IDOC-OPE: instrument operations,

IDOC-DATA: data management and data value chain: to produce the different levels of data constructed from observations of these instruments and make them available to users for ergonomic and efficient scientific interpretation (IDOC-DATA). It includes the responsibility of:

- Building access to these datasets.
- Offering the corresponding services such as catalogue management, visualization tools, software pipeline automation, etc.
- Preserving the availability and reliability of this hardware and software infrastructure, its confidentiality where applicable and its security.

(See Figure 1)

IDOC-DATA is the sole object of this certification. IDOC-DATA acts as a repository of repositories sharing the scientific

themes on which the researchers of the OSU Paris Saclay have a recognised expertise of international level.

IDOC-DATA (<https://idoc.osups.universite-paris-saclay.fr>) currently hosts 63 different datasets accessible via 13 main access interfaces. These data sets are either the result of the work of the data center members or the result of external collaborations. New datasets join our platform over time, as well as new services. Their integration is subject to a described and controlled process. The volume of data exceeds 1TB, which rises to 4TB if redundancies and backups are added.

Types of repositories hosted by IDOC-DATA :

- Domain or subject-based repository:

IDOC-DATA serves as a deposit of repositories in the Universe Sciences and particularly in the following research areas identified by the European Research Council: Astro-physics/chemistry/biology; solar system; stellar, galactic and extragalactic astronomy, planetary systems, cosmology, space science, instrumentation.

- National repository system, including governmental:

Among the datasets hosted by IDOC-DATA, some of them are national in nature. For example, the MEDOC portal is intended to disseminate all solar data from space instruments in which a French laboratory has been responsible. Therefore, IDOC-DATA acts as « National repository system ».

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

OK

Brief Description of the Repository's Designated Community.

The community addressed by IDOC-DATA includes researchers and the public educated in astrophysics.

The scientific themes are grouped by communities and identified by an acronym:

- MEDOC (Multi Experiment Data & Operation Centre for Solar Physics): Solar Physics
- AMIS: Astrophysics of Interstellar Matter
- COSMO: Cosmology
- DS2 (Stellar Systems Data): Stellar Physics
- PSUP (IAS Planetary SURfaces Portal): Solar System (planets, comets, asteroids...)

Each of the themes is organized in an adhoc manner: a theme can have its own Directorial Committee (eg. MEDOC) or can have a simpler structure. Nevertheless, each theme has one scientific coordinator with respect to IDOC-DATA. The scientific coordinator acts as representative of his or her scientific field. He is an expert in his field and has an overall view

of the domain, projects, problems, and advises both IDOC-DATA and its requesters.

Reviewer Entry

Reviewer 1

Comments:

It is now clear that the object to be certified is IDOC-DATA.

Reviewer 2

Comments:

OK

Level of Curation Performed. Select all relevant types from:

C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

ok

Comments

Depending on the dataset, the level of curation may vary from C to D: given the very different origins of data, of specialized communities of the repository, the different contracts for making the data available, the implication of the Institute in the data processing and their date of entry, the deposit disseminates data with both levels of curation.

A document is built with the data provider before hosting and distributing a new dataset. It serves as a reference for the implementation of a new data service at IDOC-DATA. It explains and details the inputs needed for building and delivering this dataset service.

The process involves iterative interaction between the client and the IDOC-DATA Technical leader until all questions are answered unambiguously. This document can subsequently be used as a Data Management Plan as it contains all the elements. It will then be implemented, and the service finally given to the customer. During this process the appropriate curation is determined and applied in order to meet the ergonomic standards of IDOC-DATA portals.

It is mainly presented in the form of points to be described and questions to be answered. It also determines which datasets will be made available and to which community.

As a result of this process, some data sets will be accompanied, in addition to their description, advice for use and

selection, and access interfaces, by the codes and software tools necessary for their correct interpretation. Moreover, IDOC-DATA can also provide platforms for on-the-fly calculation.

Note: some low-level instrumental data used to build the datasets to be disseminated are hosted by IDOC-INSTR for use by the instrumental teams. They are not intended to be disseminated to the scientific communities because they would be unintelligible and can in this case be kept with a basic degree of curation (commonly B).

These data are not part of the scope of the certification.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

ok

Insource/Outsource Partners. If applicable, please list them.

IDOC-DATA is hosted by Institut d'Astrophysique Spatiale (UMR 8760, a Joint Unit of the Centre National de la Recherche Scientifique (CNRS) and of Paris-Saclay University, which is located at Paris-Saclay University, Orsay Campus).

As far as infrastructure is concerned, IDOC-DATA's internet access provider is RENATER via the services of Paris-Saclay University. The machine rooms are one of IAS's own, a second common to the Virtualdata group to which IDOC-DATA belongs, and the third proposed by Paris-Saclay University.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

OK

Summary of Significant Changes Since Last Application (if applicable).

NA

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:
OK

Other Relevant Information.

IDOC-DATA is funded by The French National Center for Scientific Research (French: Centre National de la Recherche Scientifique, CNRS) the largest fundamental science agency in Europe that organises the whole French research.

Its other direct founder is Paris-Saclay University (formerly Paris-Sud University until January 2020)

It also has strong relationships with the CNES (Centre National d'Etudes Spatiales), the government agency responsible for shaping and implementing France's space policy in Europe, which contracts and funds IDOC-DATA for data center projects related to space missions.

Depending on the datasets, the partnerships used to build them may be under the aegis of the space agencies that originated the satellite missions or may come from laboratories or organizations that are partners of these missions. These agencies, in addition to the CNES, are also ESA, NASA, JAXA, ...

The consortia can include a few laboratories or dozens depending on the size and complexity of the missions.

All datasets hosted by IDOC-DATA are derivatives of instrument products (mainly on board of space probes). These instruments are developed for scientific research and a deeper understanding of the observed astronomical objects. Derived datasets built from these initial "raw" data are the result of scientific work aimed at highlighting the characteristics of these objects or at providing access to a scientific community from another field of research. Most of the hosted research datasets are the result of international cooperation. IDOC-DATA can be the only host, the main host or a host allowing better access or redundancy of data, acting as an international repository.

More detailed information can be found in IDOC 001 Executive Summary or within the overall presentation (All documents referenced here can be found at <https://idoc.osups.universite-paris-saclay.fr/Resources/Documentation>)

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:
OK

ORGANIZATIONAL INFRASTRUCTURE

1. 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
accept

Response:

IDOC-DATA is the Data Centre of the OSU Paris-Saclay. An OSU is a perennial structure in charge precisely of observation services.

IDOC-DATA is at the heart of all the actions of the OSU for the benefit of the data, whether in the short, medium or long term.

IDOC-DATA is labelled as a CNRS " regional center of expertise" and "platform" and labelled by the CNES as a "centre d'archivage long terme" (long-term archiving center).

A CNRS platform is defined as follows: These sets of technical means operated within certain laboratories are open to the outside world, allowing a wider use of their resources, including in some cases outside the circle of academic structures. These platforms bring together equipment, software or human resources with high-level technical expertise to support research activities.

IDOC-DATA is recognised as mastering these three types of activity:

- the production, distribution and maintenance of software;
- the production of large surveys, observation or simulation data;
- the development of data archiving and distribution tools and of tools for interrogating and manipulating large heterogeneous databases;

Within IDOC-DATA, the data comes from large astronomical observatories, mainly space-based, providing large volumes of data that are made public after a short period of exclusivity. This rapid dissemination to the wider community is intended to maximise the scientific return on heavy investments.

Actions derived from the main activities are implemented in IDOC-DATA such as data processing, archiving and dissemination within dedicated structures that have the necessary expertise and resources.

To enhance the value of the observations, IDOC-DATA can provide reference data from theoretical calculations, experiments or simulations, as well as the tools needed to exploit them. The IDOC-DATA offer can also include access to reference numerical codes.

It should also be noted that during the regular validation processes by the agencies (CNES, ESA, NASA, JAXA, etc.) of the progress of IDOC-DATA projects and its ability to disseminate and perpetuate the resulting data, many of the points mentioned in this document and described below are subject to regular audits.

Annex:

Please read here IDOC-DATA more detailed commitments, missions, references:

<https://idoc.osups.universite-paris-saclay.fr/About/Missions>

Reviewer Entry

Reviewer 1

Comments:

This answer is still very long. I miss the two key sentences in this context that I do see on the website: "The overall objective of IDOC is to support the generation of instruments and then to enable them to produce scientific data. IDOC must then provide an environment that facilitates the appraisal of these data, their curation, wide reuse, mid & long-term preservation. "

Very good to see the formal letters of support also on the website!

Reviewer 2

Comments:

OK

2. 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

Compliance Level 4 suggested, there is a proper description now.

Response:

Data access is mainly open and free of charge for IDOC-DATA datasets, services and codes.

The general license for the data, unless otherwise specified in the dataset-specific copyright (according to french government recommendations <https://www.data.gouv.fr/en/licences>) is the following:

<https://www.etalab.gouv.fr/wp-content/uploads/2018/11/open-licence.pdf>

The general license for available software codes, unless otherwise specified in the specific download page is the following:

<https://spdx.org/licenses/GPL-3.0-or-later.html#licenseText>

Note: A small part of the hosted data (currently 5 among 63) may be subject to a temporary embargo or moratorium = delay before open access to give scientists time to publish.

In the French law about “open science”, this delay is under a maximum of 6 month for science, technical and medical (STM) disciplines, after scientific qualification of data. (Article 30 - THE FRENCH LAW FOR A DIGITAL REPUBLIC - OCTOBRE 2016)

This moratorium depends mainly on the contracts proposed by the space agencies concerning the data from the missions they have financed.

- As far as NASA is concerned, the data are immediately open.

- ESA usually proposes a period of 6 months. The data life cycle of the space missions to which IDOC-DATA contributes is as follows: As soon as a slice of 6 months of data is available, it is the subject of a data package in the OAIS definition which is delivered in parallel to ESA and made available on the IDOC-DATA interfaces. As a result, a potentially large part of the dataset is available, making almost 99% of IDOC-DATA data publicly accessible.

IDOC-DATA puts in place the elements to guarantee the moratorium period, and then makes this information accessible to all.

During the period of ownership, either the means of access given are concealed from the visitor without the necessary authorisations, or explicit signs describe the restricted nature of this access.

An example of an interface request authentication is shown in Figure 2

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
OK

3. Continuity of access

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

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
accept

Response:

IDOC-DATA has full support from the two organisations in charge, CNRS and Paris-Saclay University, first French university in the Shanghai ranking, which both provide funding and staff.

IDOC-DATA is recognized as a data platform by CNRS

IDOC-DATA is recognized as a long-term archive by CNES, the French space agency

IDOC-DATA's two supervisory authorities depend on the Ministry of Research, which has made commitments in terms of open science:

The CNRS and the Université Paris Saclay are equally strongly committed to this process.

In addition, CNES, the French national space agency, has accredited IDOC-DATA for data archiving missions that have been the subject of regional or national archiving agreements.

In this environment, IDOC-DATA's actions allow it to preserve and make available the original data it has collected since its creation in 2003, as well as the metadata necessary for the evaluation and reuse of data. Since then, access to the data has been ensured, with a significant evolution of the access methods that follows both the technological evolutions and the uses and expectations of the scientific communities concerned.

All this context tends to reassure on the sustainability of IDOC-DATA either in its current structure or if this structure should change, the willingness of the guardians and contributors of IDOC-DATA to perpetuate the actions towards the hosted data. This transfer could be conducted to a new entity or by distributing these services in existing structures and adapted (the French CINES for example, already certified by CTS).

If nevertheless, it happened that the services could not be reassigned by the trustees, the international nature of data hosted and made available by IDOC-DATA would allow to keep their access for most of them.

Thus, for the solar theme, most of the data are already available through virtual observatories and hosted in another global location (SDO data for example)

For the ESA missions, all the produced datasets stabilized by the instrument teams must be transferred to the ESA data center, which guarantees permanent access to these data (<https://www.cosmos.esa.int/web/esdc>).

For instance, regarding the planetary theme, the most important data from IDOC-DATA are regularly added to the Planetary Science Archive (PSA) of the European Space Agency (ESA).

What would be lost on the data described above would be the redundancy of the accesses associated with a degradation of the response times of these services. Some original ergonomic aspects of the interfaces today offered by IDOC-DATA could be compromised but would probably be compensated for as the interfaces of the other centers are renewed and their technology evolves.

Moreover, some of the services implemented at IDOC-DATA are the result of the designation of the center as winner of a call for tenders aiming at the implementation of services (e.g. Space Situational Awareness of the ESA). As these services are subject to a recurrent follow-up, there is no doubt that the donor would repeat its call for tender and would then designate a successor to IDOC-DATA.

As for the rest of the datasets not concerned by the contexts described above, their volumetry is less important than those already preserved today. It is likely that a call to the international data centers concerned by these themes would make it possible to take back a very large part of this data and make it available to communities. IDOC-DATA's current involvement in initiatives such as the virtual observatories makes it possible to create links that would then prove useful.

Annex:

Please read here IDOC-DATA references: <https://idoc.osups.universite-paris-saclay.fr/References>

Reviewer Entry

Reviewer 1

Comments:

The response is now much more on topic although again very long. Although the described context provides reassurance about the sustainability of IDOC holdings, I understand that there is no formal written agreement in place with the university and/or CNRS given any guarantees that they will take over the responsibility in case IDOC no longer exists. Therefore the level should remain a 3

Reviewer 2

Comments:

accept

4. 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:

The data disseminated by IDOC-DATA allow the study of the properties of objects in the universe (stars, planets, galaxies, interstellar medium for example), excluding planet earth.

Therefore, concerns on this subject related to the data themselves are not relevant because the study subjects are objects with no relationship to people.

Moreover, for public data, access to them through IDOC-DATA is made without registration of identifiers.

As recommended in the GDPR (General Data Protection Regulation - European Union), logs of anonymous connections are kept for one year for security and server protection reasons. Then this information is the subject of statistical extracts allowing the visualization of the uses of IDOC-DATA in a global way.

Finally, after this one-year period, these logs are deleted.

One of the IDOC-DATA staff is clearly identified with the function of security officer.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept.

In my opinion the types of data collected, curated, and distributed by IDOC are not concerned by disciplinary and ethical norms. You mention only GDPR regarding logs of access, which seems relevant.

5. 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:

The IDOC-DATA repository structure is one of the key components of the three Institutes grouped in the OSUPS, under the governance of the CNRS and the Paris-Saclay University. IDOC-DATA's existence is thus intrinsically linked to the continued existence of those two main national organizations which is of no question for the long-term. Paris-Saclay University is supported by the Ministry of Research and Education and is the best positioned French university in the Shanghai ranking.

The CNRS is The National Centre for Scientific Research is a multidisciplinary public research organisation under the supervision of the Ministry of Higher Education, Research and Innovation

These supervisors make five-year commitments, after examining the quality of the work provided by a group of experts (HCERES commission). The last validation by this commission took place in early 2019. Significant annual funding is derived from these commitments.

These commitments relate also to the provision of staff recruited on permanent posts, which ensures that IDOC-DATA has stable skills for the continuity of data provision.

This also validates long-term access to the university's resources (infrastructure, buildings, data centre, network, etc.) allowing IDOC-DATA's activities to continue over the coming years.

In addition, CNES, the French national space agency, has recognized IDOC-DATA for a certain number of data archiving missions which have been the subject of agreements linked to the space missions from which the data originated or specifically for archiving purposes at regional or national level.

The contractual financial provisions linked to these agreements are renewed annually with funds originating from the three organizations mentioned above. They allow for the recruitment of temporary staff and the provision of necessary equipment.

A proof of this lifetime are the endorsed responsibilities in spatial missions (currently beyond 2035) with the approval of the French (CNES), and international space agencies: the European Space Agency (ESA), the National Aeronautics and Space Administration (NASA), the Japan Aerospace Exploration Agency (JAXA), etc.

With regard to the efficient use of these resources, IDOC-DATA is structured and organized in a way that ensures its development and its functioning (also described in [IDOC-EX-001 Executive Summary]). This has been proven to be efficient and successful and these past 10 years.

The staff working at IDOC-DATA is mainly composed of civil servant permanent staff (French civil service with a status ensuring the continuity of their employment and their assignment thus ensuring stability and sustainability) engineers working directly with the scientific teams. This provides them the needed expertise and adequation between the data knowledge and the needs of the users. Being software engineers, they naturally have their own careers including training and professional development.

A minority of the team is composed of temporary usually (3 years duration) contractors. Their number varies with time, project needs and funding, which, although never guaranteed, is enough to ensure the needed continuity.

Our three supervisory bodies allocate funds to organize staff meetings at all levels: professional networks, national or regional workshops or conferences.

The two IAS governing bodies (CNRS and Paris-Saclay University) as well as the CNES organise training courses which staff are encouraged to register to. The space agencies also contribute to and make available training sessions in the areas of data dissemination and preservation. It should be noted that the annual evaluation of staff involves a mandatory review of these training needs, allowing for the creation of new topics if necessary.

To help the team to organize its knowledge and activities, appropriate tools are implemented to this end (git repositories, redmine projects, wikis, ...).

Because of its involvement in space missions and associated technologies, IDOC-DATA is an actor of the ECSS rules (European Cooperation for Space Standardisation) and follows them in the development and implementation of its projects.

IDOC-DATA also follows the recommendations of the CNES in the form of a Standard Reference Frame (RNC).

Synthetic and detailed organigrams are available in annex (figure 3 and figure 4) or IDOC website

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

6. 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 stated previously, IDOC-DATA staff is composed of engineers deeply participating in the research of the scientific teams building the datasets. They thus have the necessary knowledge to accommodate the data evolutions. Most of the time, they even are the dynamic of these evolutions. Moreover, the IDOC-DATA organization largely promotes the technological monitoring to explore, in advance, future technical developments that seem promising for its present and future requirements. This allows the staff members to be prepared to integrate the right tools at the right time, if and when their benefits are validated. The participation of IDOC-DATA members in regional and national networks allows reinforcing the effectiveness of this technology monitoring.

For each of the five scientific themes it hosts (solar physics, interstellar medium physics, cosmology, stellar physics, planetary surfaces, and other thematic of GEOPS and AIM), IDOC-DATA includes a scientific leader. Each of those leaders is a recognized senior scientist who acts as an adviser irrespective of each specific delimited project requirements.

As already stated in R5 a group of independent experts validates the technical and scientific orientations and the quality of the actions undertaken.

During the definition phase of each project, the leader is usually building a team of (national or international) experts in the given theme who will help in defining the requirements and act as beta testers. After the service is online, the leader presents the datasets or tools in conferences and workshops and collects feedback.

This feedback leads to changes in the interface or integrates its FAQs. A "contacts and credits" page is always available, which also allows contributors to be thanked.

The experts team also act as reviewers periodically to ensure that the tools and datasets evolve in phase with the community needs. Further communication is done through interface sites, exchange forums or the implementation of collaborative tools.

In practice, each of the interfaces giving access to the 63 datasets distributed by IDOC-DATA allows:

- to find the first level of help for the use of these data
- to contact the experts of this interface
- to find the DOI of the dataset and the contact information of the creator(s) of this dataset
- to participate in a possible collaborative exchange place.

IDOC-DATA also organises workshops to help understand the most complex aspects or new categories of data.

At last, IDOC-DATA governance is assured by the IDOC-DATA steering committee. It is designated by the OSUPS Governing Board which gives its recommendations. The IDOC-DATA steering committee nominates both the IDOC-DATA technical and scientific leaders.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity 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:

First of all, IDOC-DATA and the data it hosts must be put into context: the production of these data is carried out by instruments on board satellites or probes managed by national or international space agencies. These agencies have put in place extremely rigorous tools and procedures to preserve both the integrity of the instruments and the integrity of the data produced, whether these data are scientific or related to the control of the technical behaviour of the instruments and satellite platforms.

Concepts such as checksum validation and quality control of the acquisition context are, for example, intrinsic to the transmission of this type of data.

When delivered to IDOC-DATA, this data is subject to multiple processes to make it intelligible to the wider scientific community. These sets of actions, known as the pipeline, are coded in accordance with the strictest development standards and are subject to extensive testing to validate the best respect of the data in the measurement of knowledge. Their implementation and operation are subject to ongoing quality controls. Reliability indicators are present in the metadata on a large number of data, allowing the user to understand whether he needs to pay particular attention to their interpretation (for example, in the case of acquisition with degraded sensors).

Note: Pipeline configuration is covered in IDOC-INS-005 IDOC-OPE Instructions for Ground Segments.

The pipeline processes are all fully automated, ensuring that when data are ingested, metadata are created at the same time. All operations are logged within dedicated files.

For some datasets, tools for validating the completeness and conformity to the dataset formatting standard are implemented and condition the availability of the data.

The most commonly used formats are FITS (solar and astrophysical data) and PDS (planetary surface data). For the latter, as an example, an ESA tool (PVV) validates the completeness of the information in the datasets created.

Most of the information related to data and metadata are collected during the implementation the process of accommodating new data sets. These processes are described in within a document hierarchy. See Figure 5

Data transfers to IDOC-DATA input, internal data transfers, or transmission to other data warehouses are always performed using tools for transmission validation. The same applies to tape archiving. The origin and destination of these transfer are always identified clearly as partners systems and are the result of fully described processes.

As an example, IDOC-DATA act as a backup repository for the CDS VO services (CDS is already CoreTrusSeal certified - <https://cds.u-strasbg.fr/CTS-CDS>). The data transfers involved in this collaboration are described both from the point of view of the machines and networks involved and from the point of view of the organization and persons responsible.

Active storage makes use of distributed storage technologies (CEPH technology) which also guarantee that the data are not modified. CEPH periodically scrubs placement groups to ensure that they contain the same information. Low-level or deep scrubbing reads the object data in each replica of the placement group to ensure that the data is identical across replicas.

All instances and persons involved in the production and management of IDOC-DATA data are fully involved in validating the relevance of the metadata associated with the dataset. The document IDOC-INS-004 IDOC-DATA Instructions for Data Ingestion and Curation covers these processes. Particular care is taken to ensure that the keywords used are both relevant, unambiguous, and allow an appropriate and efficient path through the dataset.

The adequacy of the metadata allowing the best ergonomics of the access interfaces is a recurrent subject of the users' committee and their recommendations are carefully taken into account.

Here, too, the space agencies make their support conditional on IDOC-DATA providing evidence that the data have been properly monitored and that their integrity is guaranteed. As proof of this attention, IDOC-DATA must use data validation tools and demonstrate at regular intervals that all control points are fully complied with. For example, validation using the PVV tool (description here: <https://www.cosmos.esa.int/web/psa/psa-tools>) is systematically applied to OMEGA data (IAS instrument of the Mars Express mission).

When the situation requires it, the data production pipelines are designed to identify and distinguish the elaborated versions.

Code versioning is completely handled by the development and release tools.

The instruments and their data production are published in peer-reviewed journals. These articles reflect the actions taken to ensure the quality of the information distributed.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Very thorough and detailed response.

Reviewer 2

Comments:
Accept

8. 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:

In order to integrate IDOC-DATA, a new dataset requires approval from the steering committee which will make its decision according to the usefulness for the communities gathered around IDOC-DATA. After approval, it is examined both scientifically and technically. As this external dataset might not be complete, organized, documented, or can presents other insufficiencies for long-term preservation (metadata, format...), processes are applied to remove or mitigate these defects in order to obtain the appropriate level of curation. This is done by following the document IDOC-INS-004 IDOC-DATA Instructions for Data Ingestion and Curation.

Appraisal is therefore conducted by IDOC-DATA according to general technical aspects and relevant thematic considerations.

The decision to retain a dataset over the long term is generally based on whether or not there is a more relevant or accurate dataset in the state of scientific knowledge built from the information initially acquired. Thus, for the same set of raw data from an instrument, successive versions of derived data sets may be made available depending on the improvement of the understanding of the operation of the instrument or the availability of finer ancillary data or the

improvement of algorithms.

Depending on the size of the datasets, these older versions may be retained, but in most cases the volumes involved will result in their replacement by the most complete version.

Nevertheless, the software versions that allowed their construction are kept, allowing in almost all cases, reconstruction if necessary. It must be noted that this type of demand has not emerged in the last 20 years.

The IDOC site gives practical information on each of the hosted datasets. Their respective DOIs lead to a landing page with more detailed information: each dataset is accompanied by its production context and all the information essential for scientific interpretation as well as any necessary tools and codes.

As an illustration, the following implementation of Sitools for the PSUP database give access to a detailed description of the data through the interface: psup.ias.universite-paris-saclay.fr

As IDOC-DATA aims to make available as much of its data hosted in virtual observatories as possible, it complies with their recommendations and expectations. In these rules are of course a whole set of recommendations related to the maintenance of data and metadata. In the latter case in particular, IDOC-DATA relies on the results of the IVOA working groups for the naming of metadata keywords and their use. IDOC-DATA also follows the recommendations and work of the RDA (Resource Data Alliance). For some datasets and formats, the respect of keywords is an essential part of the validation of the dataset by space agencies (PDS format for example).

In R14 are detailed some of the data formats authorized at IDOC-DATA

With respect to metadata, when considering the integration of a new dataset into IDOC-DATA, the adequacy of the metadata and the objectives for the useful distribution of the data must be verified. It has already happened that, with the agreement of the data producers, metadata have been added, for example for:

creating metadata from the combination of existing metadata to enable more efficient indexing of the dataset in the metadata base.

making a metadata more readable (e.g. convert a given time in seconds from an arbitrary date to a humanly identifiable time value)

As already mentioned in other points of this document, space agencies are attentive to the cost-effectiveness of the data produced by their very high-cost space missions. They are putting in place various actions to audit this cost-effectiveness. IDOC-DATA ensures that it fulfils its commitments with regard to these actions. In concrete terms, this can take the form of stages in the development of ground segments where the production of data is confronted with requirements that reflect scientific expectations. For example, for Euclid's "Sciences Challenges", IDOC-DATA must provide catalogues of sources that distinguish between stars and galaxies with an error rate of less than 1%.

As IDOC-DATA is funded to be a long term archive, no plan has been made for removing items.

Reviewer Entry

Reviewer 1

Comments:

Accept.

Document 004 provides a great overview.

Reviewer 2

Comments:

Accept

9. 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:

The processes aimed at integrating new data repositories into IDOC-DATA, making them available to user communities and the means implemented to ensure their sustainability are described in the documents in the documents already introduced in R7.

Of course, in order to complete the picture, it is necessary to add to it IDOC-OD-002 IDOC Risk analysis and management.

This document describes the risks identified and the means to mitigate them for the following areas:

- Risks to the very existence of the structure

- Organizational risks

- Infrastructure risks

- Physical network links risks

- Network access risks

- Procedural, People & System risks

- Risks related to infrastructure design.

- Data Risks

- Processing Risks

- Software design & development risks

- Operational risks

- Technological risks

Other elements contribute to the stability of the infrastructures deployed in this framework.

We can name the systematic use of Redmine to trace projects and actions of implementation and the equally consistent use of GitLab for the efficient monitoring of all developments.

There is also its counterpart for the staff in charge of the infrastructure allowing them to put each element of the infrastructure in its operational context. There are also valuable guidelines for use, updated as interventions are carried out, allowing interventions to be completed safely and fully (especially during urgent interventions outside normal working hours).

IDOC-DATA has adequate control of the 3 infrastructure plants and their associated networks which ensure the sustainability of the data by multiple redundancy actions and hierarchical backups.

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
Accept.

10. Preservation plan

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
This is at Compliance Level 3, based on current information.

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The IDOC-DATA repository, due to its labeling as a CNRS Regional Center of Expertise since 2013 and its recognition as a "long-term archiving center" by CNES since 2016, assumes the responsibilities for the long-term preservation of datasets. To do so, IDOC-DATA relies on a clear strategy that describes the objectives, means and actions required for data preservation. See Figure 6.

At IDOC-DATA preservation of digital data over time pursues four main objectives:

- Preserve the information

- Preserve intelligibility.

- Make it accessible

- Make it usable in a scientific way

These four objectives aim to perpetuate not only the data as such but above all their capacity to be used effectively by the user communities.

Ensuring that all four objectives are met means that it is necessary to validate over time that the infrastructures, tools, interfaces, descriptions, etc., which are the environment of the data and allow its use, retain their relevance for the understanding and effective use of the data.

IDOC-DATA's strategy in this field is fully described in the reference document (DOC-0D-006 Instructions for Data Preservation document), Applying this strategy identifies how to mitigate the four main risks either globally, by category of data and then at the level of a particular dataset.

This helps to determine the specific points of attention for that dataset that will be added to the global points of attention that IDOC-DATA knows to pay attention to.

Over time, these points of attention are validated in a cyclic way, and this pattern maintains the intelligibility of the data. IDOC-DATA's final objective is to ensure permanent access to the intellectual content of the data, and not to a simple flow of bytes.

The IDOC-DATA security officer also ensures that information assets are adequately protected. In an equally important part of its activity, it creates initial information and moral commitment for users which he extends with relevant and measured reminders.

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
accept

11. 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 reader will discover without ambiguity on reading IDOC-EX-001 executive summary that the strong intermeshing between the scientific communities at the heart of the IAS's activities and IDOC-DATA's staff and actions. This intertwining guarantees the scientific effectiveness of IDOC-DATA's approach and its pragmatism. Therefore, this ensures the scientific data quality.

Data quality refers to the state of qualitative or quantitative pieces of information.

Quantitative aspects:

As already mentioned, IDOC-DATA uses exclusively automatic processes for data retrieval and production. These processes are accompanied by precise and attentive monitoring, also automatic, to ensure the completeness of the information.

Example of a product table to control the data recovery from the SDO satellite: see Figure 7

The successive evolutions of the data access interfaces are intended to allow an increasingly easy access to the data made possible by the evolution of technologies. This continuous improvement of our tools ensures that information retrieved from one of our interfaces will be at least as easy to retrieve in the future. Consequently, the respect by the authors of our recommendations for referencing our sites intrinsically makes it possible to retrieve information afterwards if the context of the combination of parameters used in the initial search is provided.

Authors are encouraged to cite IDOC-DATA or its components such as here:

<http://idoc-medoc.ias.u-psud.fr/sitools/client-user/index.html?project=Medoc-Solar-Portal>

Qualitative aspects

Delivering high quality data is one of IDOC-DATA commitments.

Metadata are added to the data in order to give some information about the data quality, for instance with the psup metadata quality level: as described in the EAICD_OMEGA document, (Experimenter to Planetary Science Archive

Interface Control Document) a quality index is provided ranging from 5 to 0 depending of compression errors, tiny or big holes in the spectroscopic images, etc. The Keyword is named as DATA_QUALITY_ID and is referenced in the PDS dictionary.

The metadata themselves are checked using VO validators in case of data sent through the virtual observatory.

This site shows the results of the VO services validation through eurovo:

<http://registry.euro-vo.org/evor/#results> (please search IDOC)

This intense preoccupation of IDOC-DATA to make its datasets conform to the standards of the corresponding designated community makes the number of remarks on the data or metadata that come up through the contact pages of the interfaces extremely limited.

Depending on the characteristics of the data production, the access interfaces to these data can refer to the articles published in peer-reviewed journals that enabled their creation.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

12. 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 answer given in R9 provides general elements of an answer at this point as well. The diagram in Figure 8 describes these processes and workflows:

Of course, in addition to these general principles stated and described in the various documents of the IDOC-DATA referential, each pipeline (software package that produces or formats new data in an existing dataset) that is made for one of the IDOC-DATA repository is the subject of a description and specific documentation. This documentation permits to understand the details of the actions and software implemented, the automatisms deployed, the procedures for disaster recovery, ...

These sets are materialized in documents known as “Data Processing Agreement” that are approved by the parties involved (e.g. IDOC-DATA and a space agency)

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
Accept

13. 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:

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:

Data discovery processes are fully mastered by IDOC-DATA. The document IDOC-INS-006 IDOC-DATA Instructions for Data Preservation is devoted to the description of IDOC-DATA's strategies, technologies and ambitions in this field.

Data dissemination and identification at IDOC-DATA are based on the FAIR principles (especially the findability, accessibility and interoperability, as the reusability principle is more one principle of the preservation section).

The IDOC-DATA platform has decided on a nomenclature for the management of the DOIs assigned under its aegis (via DATAcite):

The common prefix is 10.48326.

10.48326/IDOC has been defined for IDOC-DATA as a platform (thus to identify for example the IDOC website) then this first DOI is broken down according to the thematic areas hosted by IDOC-DATA.

Here as an example, regarding data sets or information from planetology:

10.48326/IDOC.PSUP

Then these elements are declined again by data set (we will indicate these DOIs in the references of the interfaces).

example:

10.48326/IDOC.PSUP.OMEGA for OMEGA instrument data of the Mars-express probe.

Note: the level of declination used is not limited if this hierarchy created allows to finely account for the structure of the concerned interface, data or code:

As an example, the information on Oxia comes from two CTX and HIRISE instruments hosted on the MARSSI platform of the PSUP portal.

This gives two DOIs to be used depending on the origin of the data used:

<https://doi.org/10.48326/idoc.psup.marssi.ctx.oxia>

and

<https://doi.org/10.48326/idoc.psup.marssi.hirise.oxia>

DOI's have been affixed on the different datasets to allow their referencing.

Codes and some other tools have also been referenced by DOIs.

As stated in reply to R0 and R1, the ergonomic provision of data is one of IDOC-DATA's major activities.

In particular, the reader will discover that "REGARDS", the successor to the current data provision framework, has integrated the OAIS model into its development, which will allow a more integrated and automated follow-up of all the associated recommendations. REGARDS is to be implemented at IDOC-DATA in the end of 2022 and the access

interfaces will then be gradually migrated to this new architecture. REGARDS framework is developed for all of its interfaces by the national french spatial agency (CNES). IDOC-DATA, as privileged partner of the CNES on these subjects, has contributed to the REGARDS objectives and architecture. (RegardsOss <https://regardsoss.github.io/>)

The transition to a hierarchy of DOIs adapted to the granularity of the use of the data sets is under evaluation.

IDOC-DATA is closely following the work of the RDA granularity working group:

<https://www.rd-alliance.org/groups/data-granularity-wg/> Here again, the arrival of REGARDS will allow a more integrated management of the automatic generation of DOIs associated with the dataset.

On another aspect of quotations, IDOC-DATA encourages its users to mention the use of its services as for example here: <https://idoc.osups.universite-paris-saclay.fr/policy> or more specifically in each thematic interface.

Moreover, the IDOC site (<https://idoc.osups.universite-paris-saclay.fr/>) is used as a landing page for some of the thematic portals; afterwards, each of these portals gives access to specific and adapted information and data search actions.

IDOC-DATA, in accordance with the recommendations of its supervisors, implements the conditions for open science, which means in its field, contributions to virtual observatories and their recommendations (IVOA, EuroVO, ..): most of data distributed by IDOC-DATA are available via machine-to-machine queries or through VO queries.

All the DOIs which are pointing to IDOC-DATA are referenced here: <https://search.datacite.org/repositories/inist.idoc>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

14. 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:

Data reuse is the main objective for all data hosted in the IDOC-DATA repository. Answers on how it is done can be found mainly in document IDOC-INS-004 IDOC-DATA Instructions for Data Ingestion and Curation and IDOC-INS-006 IDOC-DATA Instructions for Data Preservation.

Beyond metadata, IDOC-DATA sets up the appropriate environment for the data. The now long-standing practice at IDOC-DATA of formats such as PDS has given the team a strong awareness of these dimensions. The PDS repositories we build inherently embody the notion of data environment of all the information useful to their understanding.

Concerning the formats, these are chosen to allow the objectives of dissemination, interpretation and sustainability of the data.

The most commonly existing formats are:

- FITS,

- JPEG2000

- NETCDF

- PDS3&4

As an example of the strategy described above, IDOC-DATA has encouraged the scientists concerned to migrate the data produced in .sav format - a proprietary format linked to the IDL language - to NETCDF format. (NetCDF is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.) As it is also a community standard for sharing scientific data it offers a much better opening and reduces both the financial and sustainability risk compared to after-sales service. IDOC-DATA has developed tools for converting existing files and helped the communities concerned to modify their production tools.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

15. 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:

Answers to this point can be found mainly in document IDOC-OD-002 IDOC Risk analysis and management.

Since its creation, IDOC-DATA has chosen to build its hardware and software infrastructure on the most standard, open-source elements possible and seeks to limit the use of proprietary formats in all areas. This implies for instance the use of widely used open-source systems, platforms or software packages:

Linux (Debian mostly), Proxmox, Ceph, Mysql, Postgresql, Apache, ...

System and network administrators are using the Redmine project management platform (restricted user access) to manage version upgrades and evolutions.

As described in R5 these employees receive continuous training and are active in professional networks that enable them to keep pace with the latest technologies and to deduce from shared experiences those that will be most useful for IDOC-DATA.

IDOC-DATA's development plans are prepared with the help of user committees and amended and validated by the steering committees.

All the information relating to the measures to be implemented in the event of incidents or accidents (disaster plan) is described on a dedicated website (for the sole use of IDOC-DATA operators for security reasons) in the form of clear procedures to be followed by those involved.

Figure 9 provides an overview of all the areas covered by this documentation.

The figure 10 shows the different aspects covered, for example, for the "backup" theme.

In this site, each infrastructure element involved in IDOC-DATA services is described. Its implementation methodology is also detailed, as well as its maintenance actions, contact points (suppliers, experts, etc.) and incident resolution sheets describing the verification, first aid and resolution actions.

Accordingly, figure 11 shows the level of detail provided to the user to control and operate a particular infrastructure element and the recovery procedures to be followed in case of an incident.

It should be remembered that in order to preserve IDOC-DATA's business continuity as much as possible, its resources are distributed over three sites, each provided with a separate, backed-up power supply, air conditioning, firewall and network. For the applications and data for which IDOC-DATA has availability commitments, the survival of only one of these sites enables the delivery of the associated services.

IDOC-DATA uses many tools, for example :

- for transmission, creation and backups of datasets,
- for infrastructure monitoring, control and operation,
- for service monitoring,
- for the supervising of software development,

with a permanent concern for the automation of tasks in order to make them as stable as possible.

IDOC-DATA is also monitoring the interfaces usage in order to report to users and steering committees on the cost-effectiveness of the efforts invested.

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
Accept

16. 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:

Answers to this point can be found mainly in document IDOC-INS-006 IDOC-DATA Instructions for Data Preservation and IDOC-OD-002 IDOC Risk analysis and management.

The IDOC-DATA organization chart clearly highlights the position of information systems security manager, which is currently held by a person specializing in these fields (she has set up and carried out several training courses on this subject for CNRS laboratory staff). In each of our supervisory bodies, this function is linked to hierarchies of responsibility for information systems security, which ensure that national directives are taken into account.

These directives are accompanied by the deployment of tools, training, rules and recommendations for all personnel.

Of course, in the particular case of IDOC-DATA and its contractual relations with the space agencies, the special attention paid to the security of systems and networks in this technological context is extremely important to our operations.

It should also be remembered that for each of our projects, the agencies require the maintenance and permanent monitoring of risk portfolios that describe the best conditions for mitigating them.

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 CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response:

Reviewer Entry

Reviewer 1

Comments:
Thank you very much for the additional work and evidence. The application has greatly improved in the course of the iterations and I suggest to grant the CoreTrustSeal to the applicants.

Reviewer 2

Comments:
Authors have answered the concerns.