



Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository Platform for Archiving CINES (PAC) complies with the guidelines version 1 of 2010 set by the Data Seal of Approval Board. The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2010 on March 15, 2011.

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

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version:	1 June 1, 2010
Guidelines Information Booklet:	DSA-booklet_1_June2010.pdf
All Guidelines Documentation:	Documentation
Repository:	Platform for Archiving CINES (PAC)
Seal Acquiry Date:	Mar. 15, 2011
For the latest version of the awarded DSA for this repository please visit our website:	http://assessment.datasealofapproval.org/seals/
Previously Acquired Seals:	None
This repository is owned by:	CINES 950 Rue de Saint Priest 34097 Montpellier Hérault France T +33467141414 F +33467523763 E olivier.rouchon@cines.fr W http://www.cines.fr/

Assessment

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository doesn't deal with data producer directly, but contracts with entities (government institutions, university libraries, etc.) responsible for gathering and/or digitizing data for which quality has already been validated (e.g. PhD theses, publications, etc).

Before the submission, the repository assists the data producer in selecting the appropriate information which will be used as metadata; the repository also validates the quality of the data to be preserved as only scientific and technical data are eligible for preservation.

At the time of submission, metadata and data are validated against defined criteria: compliance to a generic structure of submission information package (SIP), validation of file formats, etc.

Functional and technical documentation is available at the repository

(<https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/6f5c38da-69e5-4d67-8f9c-ecd72d094534/PAC%20-%20Sp%c3%a9cification> and

<https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/3005eac0-dfbd-4d71-b936-39b28daf0e8e/PAC%20-%20Sp%c3%a9cification> to provide assistance to the data producer during the different steps of the submission (building of the information package, transfer, control, etc.)

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

A limited set of file formats has been selected by the repository to ensure full characterization of the files provided by the data producer and allow future format migrations. An online tool is also available for the repository to check the correctness of the files to archive (see FACILE).

Only data in preferred formats is accepted on the data repository, but new data formats can be added to the list of preferred file formats pending validation by an internal committee of experts.

More information :

<http://www.cines.fr/spip.php?rubrique224>

<http://www.cines.fr/spip.php?rubrique256>

<http://facile.cines.fr/>

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

A generic structure of submission information package (SIP) has been put together and combines data along with metadata to be preserved. Any transfer must comply with this structure, otherwise it is rejected. The XML schema for the SIP is available here: <http://www.cines.fr/pac/sip.xsd>

Data format required by the repository is based on a list of preferred formats (see guideline 2).

Metadata format required by the repository is based on DCMI standard.

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository has two main activities – as defined by the mandate given by the French ministry of higher education & research: high performance computing and long-term preservation of electronic documents. A law published on August 7th, 2006 also reinforces the mission of the repository, stating that it is the official preservation centre for electronic PhD theses.

The repository promotes its activities in the domain of the preservation of digital documents through various working groups and seminars.

In December 2010, PAC has been given consent by the French National Archives to preserve digital material.

See also:

<https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/105be5be-a63a-40d5-b229-096bd7ddcf62/lettre-DGRI-DGES-missions-du-C>

https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/18a55e92-edeb-4a20-be0c-ed9294535168/Arrete_du_7_aout_2006_version_

<http://www.cines.fr/spip.php?rubrique223&lang=en>

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023278648&dateTexte=&categorieLien=id>

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository is under the trusteeship of the French ministry of higher education and research. It is legally the official preservation centre for electronic PhD theses.

Any data producer who deposits on the repository is under contract. Roles and responsibilities of the parties involved (data producer, repository, data users) are defined in this contract (see

https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/7d425705-7f63-462e-9938-060b62134019/Convention_Archivage_CINES_V

The repository conditions of use – i.e. criteria for eligibility of preservation projects – are published on the repository website, the interface specifications are also available online.

Recurrent reviews are scheduled with data producers to ensure the repository conditions are complied with.

A full time archivist also ensures that the repository processes and systems comply with national and international laws.

See also: <http://www.cines.fr/spip.php?rubrique219>

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

A document – “Politique d’archivage du CINES” – summarizes the repository strategy and objectives in terms of preservation. This is the basis for the risks analysis. A risk management plan has been put in place to ensure that events which could impact the repository preservation strategy are identified and managed.

There are also documented processes and procedures in place to ensure quality management for the storage of data: authentication of users through LDAP catalog, multiple on-line copies, tape backups, file integrity checks using hashing algorithms, disaster recovery plan. Online monitoring tools (Nagios, Cacti) have been put in place to check application and storage availability.

An initiative is being envisioned to replicate data in the BnF (French national library)

Risk management plan:

https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/bd7147e7-a328-4a8f-be13-bbbcf8b11575/PAC_Plan_Gestion_Risques_v6.pdf

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The “Politique d’archivage du CINES” document, which summarizes the preservation strategy of the repository can be accessed here :

<https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/2422e254-ff26-4ffa-a7b7-c37498987343/CINES-DAD-PA.pdf>

A risk management plan has been put in place to identify and mitigate risks which could impact the ability of the repository to guarantee durable archiving.

A preservation plan is also being put in place to document technology watch, evaluate obsolete and emerging file formats, define physical & logical migrations procedures, etc.

The business processes for file format technical watching (obsolescence, etc) and file format migration are available here :

<https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/21035109-9e6b-4845-a16b-7f3a0dbdc660/Expertise%20formats.pdf>

<https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/5feab7d9-fe4e-4b34-bb54-bc11df55e1f2/Migration%20logique.pdf>

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

Any data transferred has to comply with the following criterions: the SIP is well-formed and sticks to the structure defined by the repository; all files provided are compliant with the specifications of their format.

All workflows, selection process and document lifecycle are documented in the functional specifications of the repository (see

<https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/6f5c38da-69e5-4d67-8f9c-ecd72d094534/PAC%20-%20Sp%20c3%a9cification%20p17-19>).

Decision making processes have been documented for the file format expertise and watching, as well as for file format conversion, data access, etc. (see

https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/8353d8d2-0757-4a26-99cc-1eb2a2b4f500/PAC_Access_Process_Flow.pdf for an example for the access process).

See also <http://www.cines.fr/spip.php?rubrique244&lang=en> for a description of the transfer process between the data producer and the data repository

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

Any data producer who deposits on the repository has a contractual agreement signed with the repository, in which roles and responsibilities of the parties involved (data producer, repository) as well as the data user community are defined. This agreement includes the nomination of an executive board in charge of periodic reviews & reports. A sample of the contract can be found here:

https://alfresco.cines.fr/alfresco/d/a/workspace/SpacesStore/7d425705-7f63-462e-9938-060b62134019/Convention_Archivage_CINES_V

To date, the only data users accessing the data repository are the data producers themselves – making the repository a dark archive.

Reviewer Entry

Accept or Reject:

Accept

Comments:

10. The data repository enables the users to utilize the research data and refer to them.

Minimum Required Statement of Compliance:

2. We have a theoretical concept.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The data repository provides access tools (search engine, etc) to browse the catalog of archived data and request copies of data sets.

The data provided is either the original version or the latest version of the archived documents (in the event migrations have occurred).

A search engine allows data users to find data by querying against the 50 metadata qualifying a document (see screenshots at

https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/49335285-e633-45b0-80f6-7e4ba896dbd3/PAC_webtool_screenshot.pdf).

Each data set has a unique identifier based on ARK so that it can be referenced in other publications or web pages.

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository stores the electronic signature (SHA-256) of any data stored, and periodically checks that this signature remains the same. This applies to the files containing the metadata as well.

The ingest process also allows the data producer to provide an electronic signature (MD5, SHA-256) for each documents transferred on the repository. The repository will compute a new checksum with the same algorithm and compare it to the initial one to detect potential corruption during the transfer.

Availability is monitored using supervision tools (Nagios, Cacti).

The set of metadata defined to qualify documents includes version numbers to allow multiple versions of the data.

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

Authenticity and integrity policies are described in the contract between the data producer and the repository. When migrating data to an emerging format, links to the previous versions are maintained. Any operation on the data (metadata updates, migration, access, etc.) is logged. All repository logs are self-archived in the repository. Versioning of data is permitted using metadata and unique identifiers provided by the repository, as well as relations between datasets. Relations are controlled after the submission to ensure the referred dataset does exist.

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline can be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The technical infrastructure complies with the OAIS model – it is made of three logical servers (ingest, storage, access) on which the different functional entities are deployed.

See also: <http://www.cines.fr/spip.php?rubrique244>

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

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

Evidence:

The repository access module can be accessed from the following link :

<http://pac.cines.fr:9000/webarcsys/001-frmIdent.jsps>

Thought, there are security restrictions based on the IP address; in order to get access to the link above, auditors should contact the repository administrator.

Screenshots of the access module can be accessed here :

<https://alfresco.cines.fr/alfresco/n/showDocDetails/workspace/SpacesStore/49335285-e633-45b0-80f6-7e4ba896dbd3>

At present, the data producers have their own platform to allow access to the information to users. Thus, the access module on the repository is only used by the data producers to get the archived copy of their data document should they lose the one they had on their own platform – the repository is considered a dark archive.

In any case, the contract between the data producers and the data repository defines the community and the rules to deposit and access data. The access module will conform to this agreement.

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

0. Not Applicable.

Evidence:

All data consumers are authenticated when accessing the repository catalog and have accepted the policy of data use as a prerequisite to get an account to browse it.

Reviewer Entry

Accept or Reject:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

0. Not Applicable.

Evidence:

At present, the data producers have their own platform to allow access to the information to users. Thus, the access module on the CINES repository is only used by the data producers to get the archived copy of their data document should they lose the one they had.

Reviewer Entry

Accept or Reject:

Accept

Comments: