CENTRE NATIONAL D'ETUDES SPATIALES

SGDS

Direction du Centre de Toulouse
Produits et Segments sol
Techniques de Valorisation des données et Ingénierie sol

Edition : 01 Date : 10/06/2009
Révision : 00 Date : 10/06/2009
Réf. : DCT/PS/TVI - 2009.9421
Code diffusion : E

DIGITAL REPOSITORY CERTIFICATION CRITERIA

ANALYSIS REPORT

Written by: le : See the following document	
Validated by: le : 12/06/2009 BOUCON Daniele DCT/PS/TVI P.O. Richard MORENO	

Centre de Service Applications Sol & Scientifiques

Projet : Standard CCSDS de Certification d'Archives long terme & CDPP

DIGITAL REPOSITORY CERTIFICATION CRITERIA
ANALYSIS REPORT

Written by: Claude HUC	Circulation: Danièle BOUCON CNES (DCT/PS/TVI) Roger TROUILLARD AKKA (IS)
Approved by: Nicolas LORMANT	

GROUPE AKKA Technologies – Informatiques & Systèmes
SITE DE PARIS – 24-28, avenue du Général de Gaulle – 92156 SURESNES CEDEX – Tél. 01.41.44.35.35 – Fax. 01.41.44.35.36
SITE DE TOULOUSE – 6, rue Roger Camboulives – BP 13633 – 31036 TOULOUSE CEDEX 1 - Tél. 05.34.61.92.92 – Fax. 05.34.61.92.22
SITE DE LYON – Z.I. LES FAVIERES – 69380 LISSIEU - Tél. 04.72.54.99.99– Fax. 04.72.54.99.98

CHANGE PAGE

Vers.	Date	Section	Description of the change
01.0	30/05/2009		French version – Creation of the French document
01.1	01/06/2009		French version - Correction and complementary content following rereading by Danièle Boucon
1.0	12/06/2009		English version validated

CONTENTS

1 DOCUMENT PURPOSE

This document analyses the draft standard “Metrics for Digital Repository Audit and Certification". This analysis, based on the draft standard dated 30 April 2008, was completed by applying certification criteria to a scientific data archive centre, the Plasma Physics Data Centre, or CDPP (*Centre de Données de la Physique des Plasmas*).

Compared with the preliminary analysis carried out on the January 2009 version of the draft standard, the current draft standard shows significant improvements.

Each criterion was assessed twice:

- The criterion’s relevance with respect to the draft standard,
- The CDPP’s situation with respect to the criterion.

Only the first point is presented in this report. In certain cases, changes were suggested when considered justified.

The CDPP is briefly presented in section 3.

A number of observations and general proposals are found in section 4.

Section 5 contains a synoptic overview of CDPP analysis results.

An analysis of the time needed to complete this kind of certification audit is presented in section 6.

Section 7 lists the different sources of information used.

Finally, the detailed analysis of criteria is presented in section 8.

2 ABBREVIATIONS

Abbreviation	Name in full
AIP	Archival Information Package
CAS	*Composante d'activités scientifiques du CDPP* – CDPP scientific activities component
CAT	*Composante d'activités techniques du CDPP* – CDPP technical activities component
CCSDS	Consultative Committee for Space Data Systems
CDPP	*Centre de Données de la Physique des Plasmas* – Plasma Physics data centre
CU	*Comité des utilisateurs* – User committee
OAIS	Open Archival Information System
PDI	Preservation Description Information
RAC	Repository Audit and Certification (CCSDS working group)
RNC	*Référentiel Normatif du CNES* – CNES reference standards
SIP	Submission Information Package
SMC	*Système de management du CNES* – CNES management system
STAF	*Service de Transfert et d'Archivage des Fichiers* – File transfer and archive service

3 CDPP CONTEXT

The CDPP (*Centre de Données de la Physique des Plasmas* – plasma physics data centre) is the French national data centre for solar system plasmas, serving the space plasma physics community. It was created in 1998 by French space agency **CNES** (*Centre National d'Etudes Spatiales*) and the French national scientific research centre **CNRS** (*Centre National de la Recherche Scientifique*).

The CDPP is hosted by French space radiation research centre **CESR** (*Centre d'Etudes Spatiales des Rayonnements*) in Toulouse.

The CDPP's primary purpose to assure the long-term preservation of data relevant to the physics of naturally-occurring plasmas in the solar system, especially data from all-French experiments or experiments with significant French input. The data archived and accessible at the CDPP were obtained over a period of more than 30 years aboard satellites or from ground observatories. The CDPP continuously improves its data exploitation capabilities in order to facilitate detailed scientific investigations. For that end, CDPP develops tools and services facilitating data extraction, processing and analysis. The CDPP is also involved in the development of interoperability and virtual observatories.

The CDPP currently stores about 200 datasets from 70 space or ground-based experiments. Each dataset comprises a set of files containing experiment data (from a few hundred to tens of thousands of files per data set) and is accompanied by information allowing useful data to be found, ordered and used.

The certification criteria proposed in the draft standard are analysed by application to the CDPP. This is not a theoretical exercise but a life-size application in a real situation.

4 GENERAL COMMENTS AND PROPOSALS ON THE DRAFT STANDARD

The version available since 30 April 2009 contains many improvements on the January 2009 version.

4.1 DOCUMENT STRUCTURE AND ASSOCIATED TERMINOLOGY

The draft standard refers to a hierarchical structure of documentary components:

- **Mission statement**: definition of the Repository's purpose,
- **Preservation strategic plan**: translation of the mission into objectives to be reached,
- **Preservation policy**: approach chosen to reach the objectives defined in the preservation strategic plan: preservation of objects for which the Repository is responsible and access policy – How is access organized? What are its main activities? And how are these activities justified?,
- **Preservation implementation plan**: archiving policy implementation plan. Identification of the various internal and external services set up by the Repository,
- **Procedure**: specification of the actions needed to implement a service. Procedures define exactly how the crucial components of the implementation plan are actually put into practice,
- **Practices**: the elementary actions that make up a procedure.

Figure 1 represents our understanding of the dependency between documentary components.

It would be clearer to say that this is an abstract documentary model that, in reality, can result in different documentary structures, a different distribution of subjects between documents, different document names etc.

The proposed definitions are not always clear (for instance use of the synonyms "goals" and "objectives" in the definition of the Preservation strategic plan), and the boundary between the Preservation strategic plan and the Preservation policy is nebulous.

This point could perhaps be clarified by showing the relationship between these documents and the activities of the OAIS functional entities responsible for their creation (see figure 2). Within the Repository:

- The "administration" entity takes all decisions. It defines, manages and supervises the Repository's operations on a day-to-day basis in accordance with its mission. It defines or applies the archiving policy. It defines implementation in general and plans future activities,
- Specific actions and procedures may be defined globally or by the functional entities involved,
- The "preservation planning" entity fine-tunes future strategies and suggests actions for the future related to known risks such as migration or the implementation of new standards.

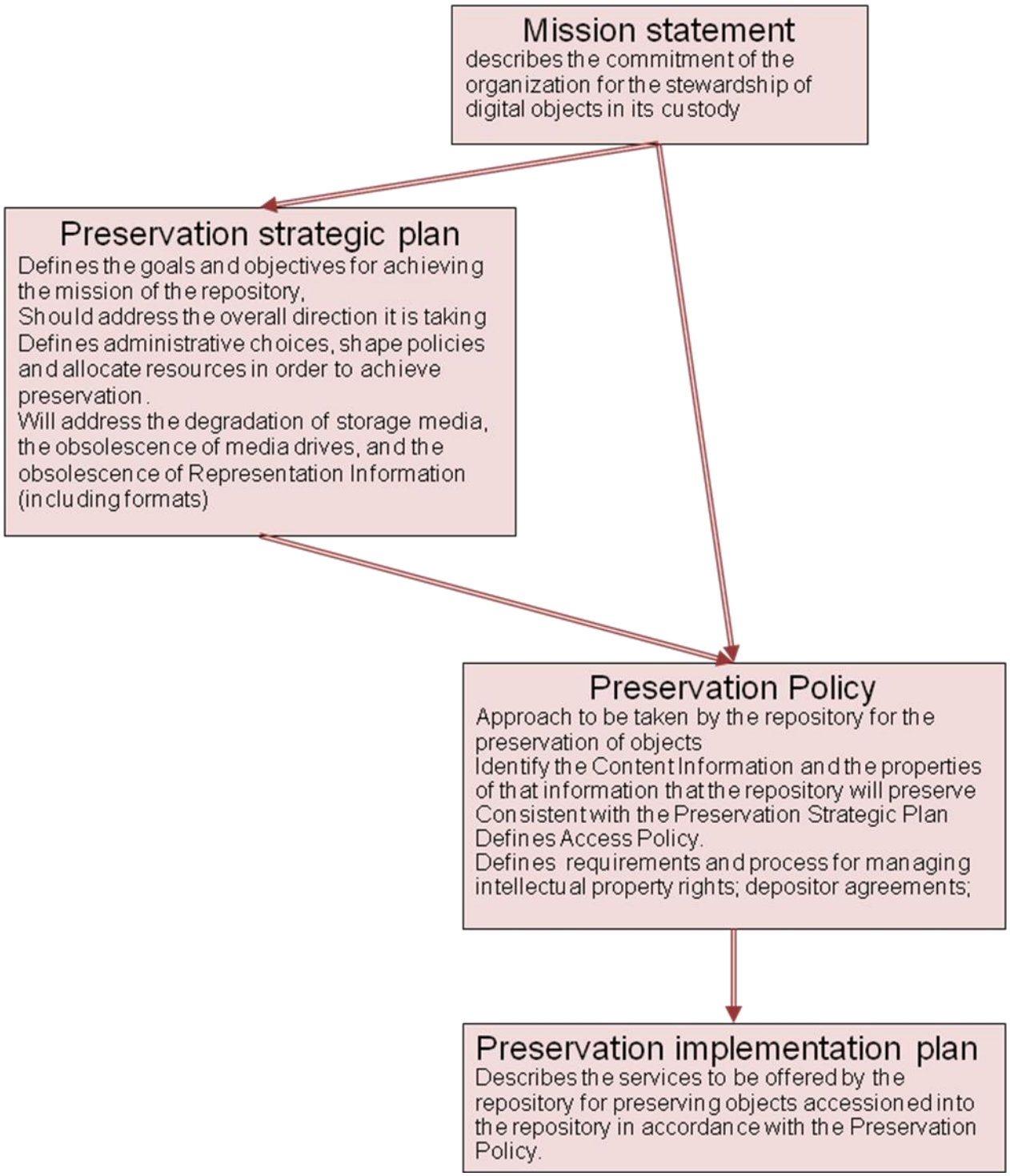

Figure 1: the baseline documentary structure

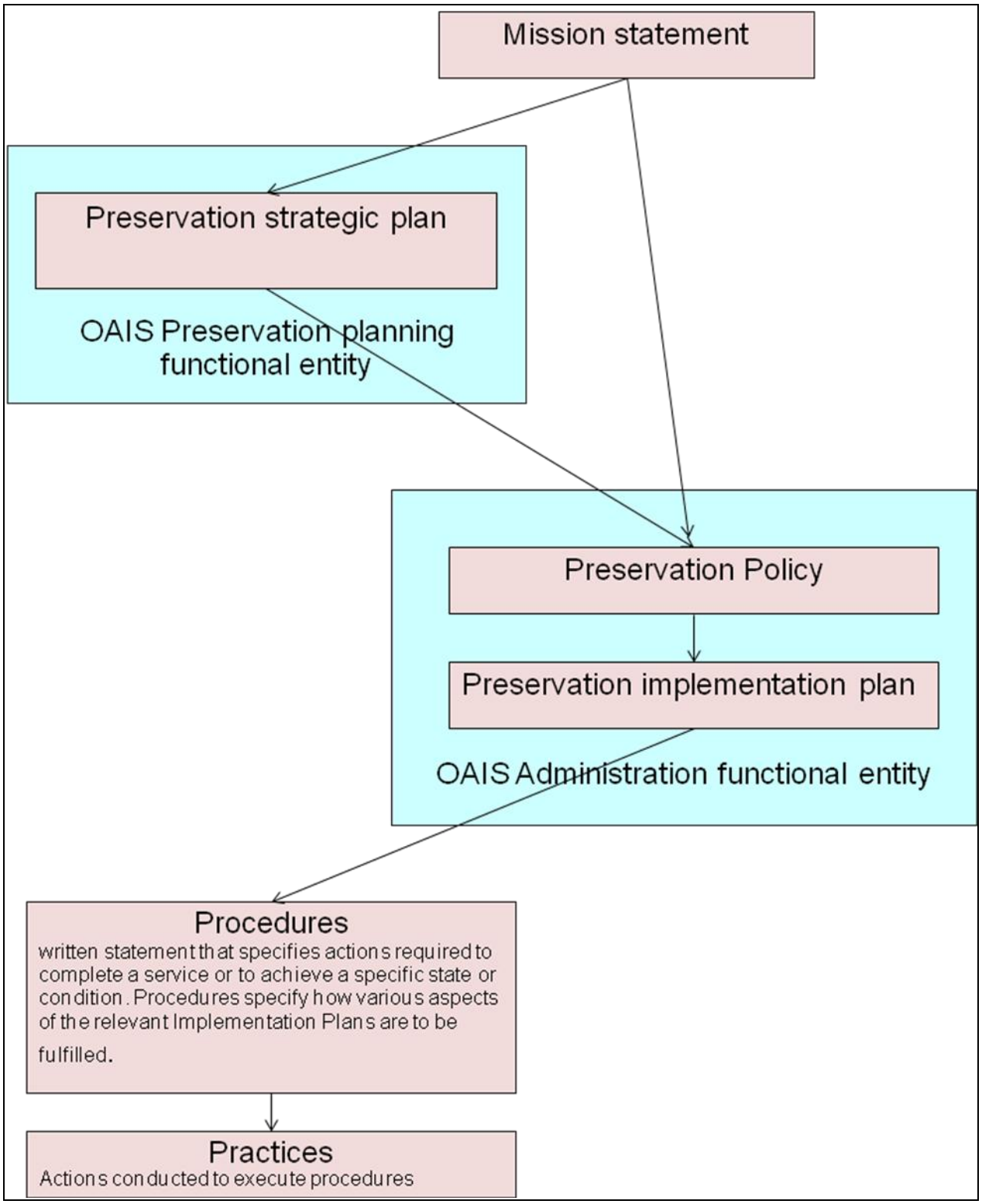

Figure 2: relations between documentary components and functional entities according to the OAIS model

4.2 CRITERIA HIERARCHY

A criteria hierarchy is relevant because it can be used to define global criteria and identify the crucial issues to be addressed as sub-criteria. However, we think that one hierarchical level should be sufficient.

Using two or even three hierarchical levels is perhaps more satisfactory from a purely logical viewpoint, but this does not facilitate the auditor's task or the determination of suitable weighted marking.

Furthermore, in certain cases, sub-criteria are linked to issues or documents from widely differing origins. The suggested changes to be made in this hierarchy are specified during criteria analysis because in a few cases, it appears to us to be useful to raise a sub-criterion up to the highest hierarchical level.

4.3 NUMBER OF CRITERIA

The introduction of a criteria hierarchy and the ability, in some cases, to process and globally assess a specific criterion and all its sub-criteria significantly reduces the number of points to be analysed and assessed.

There are many criteria and we have suggested ways of simplifying matters.

4.4 IDENTIFICATION OF CRITERIA

In the version of the draft project we analysed, each criterion has its own identifier, such as A1, A1.1, A1.2.1 and so on.

In the most recent version of the draft project, these identifiers have been removed. We believe they were very useful and in any case, auditors will have to draw up synopsis tables and determine identifiers for practical reasons.

4.5 HOMOGENEITY

There are a few inhomogeneous situations in the formulation of criteria and sub-criteria:

- Some first-level criteria address a complete subject (such as representation information, or PDI) for which several sub-criteria such as B2.5, B2.6 or B2.7 are defined to ensure the procedure is documented and applied. This appears to be useful,
- In a few other cases, the documentation aspect and the implementation aspect are separated into distinct criteria, such as B5.1 and B5.2.

We have proposed grouping or reformulating criteria in those cases identified.

We also note:

- A high level of detail which is justified for the creation of AIPs,
- A level of detail that appears insufficient for the administration field (administration according to the OAIS model), partly covered by section A. It should be possible to ask the Repository to draw up and keep an exact up-to-date description of the international standards and branch standards on which it is founded.

4.6 POSITION WITH RESPECT TO ISO 9001 CERTIFICATION

When the digital Repository being audited is part of an organization that has already been granted ISO 9001 certification, and when some of the processes defined during certification directly concern the Repository or hardware, software or the network used by the Repository, the factual elements to be analysed should be clearly differentiated from those elements that may be considered to have been already audited. However, this assumes that it was possible to check that what was previously audited was audited not only with respect to the same subjects, but also the same requirements.

We are assuming here that the documented and applied processes audited during ISO 9001 certification are not to be checked again when certifying the Repository. This point should be elucidated in the draft standard "Requirements for bodies providing audit and certification of digital preservation management systems".

4.7 CNES REFERENCE STANDARDS

The analysis of certification criteria must also take into account an in-house CNES standard: "*Data engineering and preservation - Rules and recommendations applicable to data archiving services"*. This document, written in late 2007, was based on previous work on certification criteria then tailored to CNES. The rules defined in this document have been quoted when we consider them useful.

4.8 INTELLIGIBILITY OF ARCHIVED INFORMATION

The more complex the information archived, the more difficult it is to audit. This is perfectly illustrated by the work conducted for Europe's CASPAR project (http://www.casparpreserves.eu/).

The auditor must take a particularly close look at criterion B1.7 (The Repository shall ensure that Content Information of the AIPs is understandable for their Designated Community at the time of creation of the AIP).

4.9 ANALYSIS OF PROPOSED CRITERIA

We shall refer to the exhaustive examination of all the criteria.

In the present document, we shall find:

- General comments on each of the three sections (Organizational infrastructure, Digital objects management, Infrastructure and security risk management),
- Comments on the relevance of each criterion and, in a number of cases, proposals for reformulating the criterion or changing its hierarchical position.

The results are presented as follows:

B1.3 (criterion identifier)
English definition of the criterion
Evaluation of the criterion *(is this criterion relevant for the draft standard?)*
Proposal for the draft standard (if needed)

A fuller document in French will also contain, for each criterion:

- An evaluation of the CDPP with respect to this criterion,
- Any recommendations on the subject.

The results are presented as follows in this French document:

B1.3 (criterion identifier)
English definition of the criterion
French translation of the English definition
Factual elements available: documents, logs, on-line demonstration etc.
Evaluation of the criterion *(is this criterion relevant for the draft standard?)*
Evaluation of the CDPP with respect to this criterion
Proposal and recommendation for the CDPP (if needed)
Proposal for the draft standard (if needed)

5 SYNOPTIC RESULT FOR THE CDPP EVALUATION

Full tables like that below are drawn up to have a synoptic overview of the analysis.

The CDPP's compliance with such and such a criterion is not just black or white. A criterion can be partly satisfied, mostly satisfied or fully satisfied.

We therefore chose a marking system from 0 to 4 with the following meanings:

- 0: criterion non compliant
- 1: criterion slightly compliant
- 2: criterion half compliant
- 3: criterion mostly compliant
- 4: criterion fully compliant

It appears to us to be both illusory and needless to seek to be more precise.

We preferred to use figures rather than letters (A,B,C,D,E) so as to establish an overall Repository mark once the marks for each criterion have been added together.

The CDPP weighting makes it possible to put the importance of one or the other criterion for the CDPP into perspective.

Criterion	Sub-criterion	Summary	CDPP mark	CDPP weighting	Weighted mark	Max. mark
A1.1		Shall have a mission statement that reflects a commitment to the long-term retention of, management of, and access to, digital information	4	1	4	4
A1.2		Etc	2	1	2	4
A1.2.1 A1.2.2		Etc	2	1	2	4
A1.3			4	1	4	4
A2.1	A2.1.1 A2.1.3		1	1	1	4
A2.1.2			1	1	1	4
A3.1			4	1	4	4
A3.2	A3.2.1		1	1	1	4
A3.2.2				0		
A3.3			4	1	4	4
A3.4			2	1	2	4
A3.5			2	0,5	1	2
A3.6			1	1	1	4
A4.1			4	1	4	4
A4.2			4	1	4	4
A4.3			4	1	4	4
A5.1	A5.1.1 A5.1.2 A5.1.3 A5.1.4		1	1	1	4
A5.2			2	1	2	4
		TOTAL			**42**	**66**

6 TIME NEEDED FOR A CERTIFICATION AUDIT

This section aims to give a few indications on how to estimate the time needed for a life-size certification audit. The tasks carried out during our analysis may be broken down as follows:

1. Get to know in detail the draft standards “Metrics for Digital Repository Audit and Certification" and “Requirements for Bodies providing Audit and Certification of digital preservation management systems"
   - ⇨ Auditors must have all the standards applicable in the field they are auditing at their fingertips, especially the certification criteria, rules on audit conduct and the main standards applicable to digital repositories (OAIS model). The time required to acquire this knowledge is a preliminary investment,
2. Get to know the main characteristics of the Repository and the institution or company hosting it.
   - ⇨ This task is an audit activity. Assuming that synopsis documents are available, at least two or three days are needed. These documents must enable the auditor to familiarise him or herself with the key characteristics of the institution hosting the digital repository. This task must also enable the auditor to become familiar with general documents peculiar to the Repository: Preservation strategic Plan, Preservation policies, Preservation implementation plan or their equivalent within the institution.
3. Analyse the general structure of the “Metrics” document and its overall logic, and examine each criterion to assess its relevance.
   - ⇨ This is a task that we undertook but that is not part of the audit. It is helpful for drafting the standard.
4. Analyse the target Repository in the light of each criterion.
   - ⇨ This is the auditor’s main task and includes:
     - Identifying criteria that are irrelevant in the audit context,
     - Identifying criteria or sub-criteria that must be examined individually,
     - Identifying documents or factual elements providing complete or partial answers,
     - Identifying people likely to provide complementary explanations,
     - Evaluating the Repository’s degree of compliance or non-compliance with respect to each criterion,

This is probably the auditor’s most time-consuming task. The bigger the institution, the more documentation there will be scattered throughout various departments. The CDPP - hosted by CNES and CNRS - provides a good example of this complexity. The technical component of CDPP, hosted by CNES, has:

- A pooled infrastructure of hardware resources supporting the Repository,
- A file storage and preservation service,
- A service in charge of information system security,

- A service in charge of physical security (access to the site, buildings etc.),
- A service in charge of software configuration management,
- A service in charge of document management,
- And so on.

Each service has its own documentary reference system applicable to its activities.

In the present case, this task took 4 to 5 weeks.

5. Draw up a general synopsis of the evaluation of these criteria
   - ⇨ Such a synopsis provides an overview of all non-compliances, the number of major non-compliances and the number of minor non-compliances. The auditor will then be able to decide whether or not:
     - the Repository can be certified,
     - the Repository can be certified on condition that it corrects the minor non-compliances identified,
     - the Repository must correct a few major non-compliances before another partial certification audit can be carried out, focusing only on the non-compliances previously identified,
     - the Repository must first change some of its practices, documentation, processes, software, etc. before requesting another certification audit.

If the criteria were analysed correctly, two or three days should be sufficient for this synopsis, which may result in the Repository providing complementary documents likely to modify the auditor's viewpoint.

6. Make fuller recommendations for the Repository
   - ⇨ In the present case, we put forward a number of recommendations to the CDPP. These recommendations mainly lie outside the scope of an audit, but always relate to non-compliances.

7 IDENTIFICATION OF THE DIFFERENT INFORMATION SOURCES NEEDED FOR THE EVALUATION

We have access to the following sources of information:

- Documents:
  - A complete list of contents for the CDPP, PLASMA and SGDS project documentation bases,
  - Technical documents relating to SIPAD-NG,
  - CNES reference documents concerning the CDPP,
- The CDPP data dictionary,
- The CDPP site offering access to public data, http://cdpp.cesr.fr/
- The non-public part of the site (access rights to DEMETER data),
- The project server,
- Interviews with internal and external players,
- Demonstrations on the server (user part, administrator part),
- Online examination of STAF contents for the CDPP,
- The data inventory,
- The Valdo server (adding value to data) containing the operation reports on the ingest processes): http://84.14.57.149:8082/valdo/
- Documents related to ISO 9001 CNES certification (CNES Management System).

8 DETAILED ANALYSIS OF CERTIFICATION CRITERIA

In this analysis, we have kept the identifiers of the different criteria (A1.1, A1.2…) used in the draft standard available on 30 April 2009. This document has been changed since, so it is possible that certain remarks are obsolete compared with the up-to-date version.

8.1 SECTION A – ORGANIZATIONAL INFRASTRUCTURE

8.1.1 A1 Governance & organizational viability

It appears to us that, in this section, the mandate or mission that Management tasks the Repository with must be more complete. On the theme of Repository governance, the Repository's position with respect to its environment must be defined by referring to the OAIS functional diagram:

Context:

- Which Projects produce digital information that requires storage?
- Who will be using this data and what is the assumed knowledge of the user community?
- What services should be provided for users?

Digital data to be preserved by the Repository, and in particular:

- What data are involved?
- How long must the data be kept?
- What, if necessary, is the process allowing data to be destroyed?

It may be added that if the mission statement is not precise enough, the Repository must draft a document that clearly specifies the mission and have it approved by Management.

A1.1
The Repository shall have a mission statement that reflects a commitment to the long-term retention of, management of, and access to, digital information.
Evaluation of the criterion: Clear, indispensable criterion but not detailed enough because the Repository's mission statement must at least specify the type of digital information objects to be preserved.
Proposal or comment to CCSDS: it appears to us that the mission statement should at least specify which Information objects are to be preserved.

A1.2
Repository shall have a Preservation Strategic Plan that defines the approach the repository will take in the long-term support of its mission.
Evaluation of the criterion: Clear and indispensable.

A1.2.1 et A1.2.2
A1.2.1 Repository shall have an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope. **A1.2.2 Repository monitors organizational environment to determine when to execute its formal succession plan, contingency plans, and/or escrow arrangements.**
Evaluation of the criterion: Clear and justified. It corresponds to condition 6 of document DA4, which states: *"The Repository shall have drawn up a succession plan":* *"This succession plan shall be implemented should the Repository have to cease operating. The data to be preserved should be transferred to another Repository. The plan shall describe how all the data and metadata shall be extracted prior to transfer".* *However, it appears to us that this contingency aspect, which relates to exceptional situations should be distinguished from a commitment to continuity, service availability and performance for both users and producers. This point is not addressed in sufficient detail.*
Proposal or comment to CCSDS: clearly address the Repository's commitments with respect to continuity, service availability and performance for both users and producers.

A1.3
Repository shall have a Collection Policy or other document that specifies the type of information to be preserved
Evaluation of the criterion: Valid and justified.

8.1.2 A2 Organizational structure & staffing

A2.1, A2.1.1 and A.2.1.3
A2.1 Repository shall have identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties. **A2.1.1 Repository shall have identified and established the duties that it needs to perform.** **A2.1.3 Repository shall have an active professional development program in place that provides staff with skills and expertise development opportunities.**
Evaluation of the criterion: Clear and justified. We think that sub-criterion A2.1.2 on resources should be separated. It is addressed further below. Document DA4 also specifies that **"**the Repository has established an organization and development plan*:"* *"The Repository shall have established an organization plan identifying resources and skills, defining the roles and responsibilities of staff and a development plan defining all the tasks needed to be performed to be able to fulfil the mission assigned to it in the light of an increasing number of data requiring storage, changes in the user community and constant technological developments".*
Proposal or comment to CCSDS: Make criterion A2.1.2 into a first-level criterion because, while it also concerns tasks to be performed, it appears that the justifications and decisional processes will not the same as for A2.1.1 and A2.1.3.

A2.1.2
Repository shall have the appropriate number of staff to support all functions and services
Evaluation of the criterion: This criterion covers the appropriateness of Repository resources to its ability to fulfil mission needs. These resources are not only human resources, and a more general requirement would suffice. Condition 4 of document DA4 proposes: • *The Repository shall have resources in sufficient quantity to be able to fulfil its mission. The Repository shall have identified and quantified the resources it requires. These include skills, human and financial resources and premises.* • *The Repository's resources shall be geared to its needs".* It should be noted that section A4 does not address the question of whether there are sufficient financial resources. This sub-criterion should be a main criterion. The documents concerning resources are distinct as regards both content and origin from documents used for the previous criterion.
Proposal or comment to CCSDS: modify the text for this criterion as follows: *"The Repository must have the resources needed to fulfil its duties and services (whether human, financial or other)".*

8.1.3 A3 Procedural accountability & Preservation Policy framework

A3.1
The Repository shall have defined its Designated Community and associated knowledge base(s) and shall have these definitions appropriately accessible.
Evaluation of the criterion: Clear and justified. *Condition 2 of document DA4 specifies that the Repository's mission statement should accurately specify:* • *Who are the users? What knowledge is the user community assumed to have?* • *What service must be provided for the users?*

A3.2 and A3.2.1
A3.2 Repository shall have Preservation Policies in place to describe how its Preservation Strategic Plan will be met. **A3.2.1 Repository shall have mechanisms for review, update, and ongoing development of its Preservation Policies as the repository grows and as technology and community practice evolve.**
Evaluation of the criterion: Clear and justified. Document DA4 proposes (planning part of preservation) • *The Repository shall have described and documented its preservation strategy.* • *The Repository shall assure a technology watch on the upgrading and emergence of new software, hardware and storage media with an impact on its activity,* • *The Repository shall look ahead to the user community's future needs.*

A3.2.2
The repository shall maintain policies that specify the nature of any legal permissions required to preserve digital content over time, and the repository shall acquire these permissions when needed.
Evaluation of the criterion: Clear, but should be inserted in A5.
Proposal or comment to CCSDS: regroup questions on rights to data in section A5.

A3.3
Repository shall have a documented history of the changes to its operations, procedures, software, and hardware.
Evaluation of the criterion: Clear and justified. *Document DA4 includes the following rule: "The Repository shall keep an up-to-date record of all changes in software, hardware or operating procedures when these changes have an impact on preservation strategy".*

A3.4
Repository commits to transparency and accountability in all actions supporting the operation and management of the repository that affect the preservation of digital content over time.
Evaluation of the criterion: Clear for accountability, less clear for transparency. Transparency for whom?

A3.5
Repository commits to defining, collecting, tracking, and appropriately providing its information integrity measurements.
Evaluation of the criterion: Valid and justified.

A3.6
Repository commits to a regular schedule of self-assessment and certification.
Evaluation of the criterion: Valid and justified.

8.1.4 A4 Financial sustainability

A4.1
Repository shall have short- and long-term business planning processes in place to sustain the repository over time.
Evaluation of the criterion: Clear and justified.

A4.2
Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.
Evaluation of the criterion: Clear and justified.

A4.3
Repository shall have ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).
Evaluation of the criterion: Clear and justified, part of any organization's good practices.

8.1.5 A5 Contracts, licenses & liabilities

Document DA4 emphasizes in condition 3:

The various legal aspects shall have been specified in writing:

"All the legal issues shall have been specified in writing. These issues concern the Repository's management of access rights to data and the responsibilities and special rights of the Repository with respect to the data as regards preservation of content, transformation of formats etc.".

The first issue to be addressed is the identification of the nature of the rights and the identification of the holder of these rights.

We can contribute the following clarifications on this subject:

Intellectual property rights

These are rights coming under French intellectual property law (http://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006069414&dateTexte=20090508)

With respect to the CDPP, these rights are:

- To do with software (moral and property rights). This issue is addressed by criterion A5.2. When the software has been developed as part of a professional activity, these rights belong to the employer,
- Copyright on other intellectual work, especially scientific publications. In the specific case of University lecturers and research workers, moral and property rights belong wholly to the authors and not their employer (see the following website, edited by the CNRS's legal affairs department http://www.sg.cnrs.fr/daj/propriete/droits/droits.htm).

Data produced by software are not concerned by intellectual property rights. They are not intellectual work. They are nonetheless protected by database rights.

Database rights

These rights are based on the European directive of 11 March 1996 on databases and the corresponding French Act of 1 July 1998.

"By database we mean a collection of works, data or other independent elements set out systematically or methodically and individually accessible by electronic or other means". This definition, found in the French Act, implies that the data sets archived at the CDPP constitute databases and that different types of protection apply:

- Database copyright: the person who selects the data and decides how they are presented is considered to be the database author and therefore benefits from copyright protection,
- Producer rights protect the investment of the person or the organization who took the initiative and risk of investment in order to constitute the database. Legislation gives this person or this organization (most often a legal entity) a right of protection similar to that of an author for his/her work. However, in this case it appears that this right belongs to the employer and not to the physical person who actually did the work. Protection expires 15 calendar years after database completion.

It appears here that any data set transferred to the CDPP by a data producer is in itself a database in legal terms. The CNRS or sometimes CNRS and CNES are the joint producers of this database and therefore hold the corresponding rights.

An analysis of all these issues should enable us to distinguish legal issues from ethics or a code of conduct applied within the scientific community.

A5.1
A5.1 Repository shall have and maintains appropriate contracts or deposit agreements for digital materials that it manages, preserves, and/or to which it provides access. **A5.1.1 Repository contracts or deposit agreements shall specify and transfer all necessary preservation rights, and those rights transferred shall be documented**. **A5.1.2 Repository shall have specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.** **A5.1.3 Repository shall have written policies that indicate when it accepts preservation responsibility for contents of each set of submitted data objects** **A5.1.4 Policies are in place to address liability and challenges to ownership/rights**
Evaluation of the criterion: Clear and indispensable. The four sub-criteria provide crucial elements for the drafting of a compliant standard agreement. Criterion A3.2.2 on the identification of rights needed to ensure preservation and transfer of these rights to the Repository should be added here.
Proposal or comment to CCSDS: Add A3.2.2 to this criterion.

A5.2
Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.
Evaluation of the criterion: Clear and justified but potentially insufficient. In Europe, rights concerning data are covered by the European directive of 11 March 1996 on databases. These rights, which concern the database producer, are distinct from intellectual property rights.
Proposal or comment to CCSDS: modify the text of the criterion to include any legal arrangements on databases.

8.2 SECTION B - DIGITAL OBJECT MANAGEMENT

8.2.1 B1 Ingest: acquisition of content

Proposal or comment to CCSDS: Apart from the proposals made for certain criteria, we think that the Repository should also be able to guarantee that the producer adheres to the commitments specified in the agreements/deposit contracts. The PAIS draft standard allows for the production of a plan of objects to be transferred. This plan can, for instance, specify that the transfer of this or the other SIP is mandatory. It can also define the number of SIPs that will be transferred, etc. In such a case, the Repository must not only ensure that each SIP transferred is complete and correct, but it must also ensure that all the SIPs expected have been transferred.

B1.1
B1.1 The Repository shall identify the Content Information and the properties of that information that the repository will preserve. • **B1.1 .1 The repository shall have a procedure(s) for identifying those properties that it will preserve.** • **B1.1.2 The repository shall have a record of the Content Information and the properties of that information that it will preserve.**
Evaluation of the criterion: Clear and indispensable. *Document DA4 specifies: "The Repository shall have clearly identified the information that is the main object of preservation".*

B1.2
Repository shall clearly specify the information that needs to be associated with specific Content Information at the time of its deposit.
Evaluation of the criterion: Clear and indispensable. The formulation is ambiguous, however, as the information required to constitute an AIP is not necessarily transferred together. *Document DA4 contains the following two rules:* • *The Repository shall have clearly identified the information that is the main object of preservation,* • *The Repository shall have clearly identified the information that must be associated with the main objects to preserve: auxiliary data, metadata, documentation etc.*

B1.3
Repository shall have specifications enabling recognition and parsing of the SIPs.
Evaluation of the criterion: Clear and indispensable.
Proposal or comment to CCSDS: regroup and hierarchize the 3 criteria B1.3, B1.4 and B1.5 as follows: ***B1.3 Repository shall have specifications of the SIPs and ingest process which verifies these specifications.*** ***B1.3.1 Repository shall have specifications enabling recognition and parsing of the SIPs.*** ***B1.3.2 Repository shall have mechanisms to appropriately verify the depositor of all materials.*** ***B1.3.3 Repository shall have an ingest process which verifies each SIP for completeness and correctness.***

B1.4
Repository shall have mechanisms to appropriately verify the depositor of all materials.
Evaluation of the criterion: Clear and justified. Document DA4 includes a similar rule: "*The Repository shall have mechanisms to authenticate the source of all data received".*
Proposal or comment to CCSDS: see B1.3

B1.5
The Repository shall have an ingest process which verifies each SIP for completeness and correctness.
Evaluation of the criterion: Valid. The SIP may be incorrect or incomplete because: • Of a transmission error—what was received is not what was sent (this hypothesis should be added to the discussion part of the criterion), • Not all the components that should be in the SIP are, • Not all the components are in the right format or not all the information expected in these components is included, • The content is invalid. Document DA4 includes a similar rule: "*The Repository shall validate each object (data, metadata…) transferred with respect to its quality, completeness and compliance with the archiving plan".* *It has two more rules:* *"The Repository shall be able to receive and manage data flows transmitted as per the volume allowed for in the archiving plan".* *And* *"The validation of formats of objects from Project applications shall be based on a formal representation of these formats".*
Proposal or comment to CCSDS: see B1.3 *If the Repository is to check that the SIP is complete and correct, then the Repository needs to have been given an exact specification of SIP structure and contents. It therefore appears useful to broaden the field of criterion B1.3 to SIP structure and contents. In the previous criteria, SIP contents must only be identified. We think that the contents should be specified.* This criterion should also specify that it is the validation of the SIP structure that is of interest, i.e. validation of the internal structure of data formats and not the validity of SIP content. "Correctness" could be replaced by "compliance".

B1.6
The Repository shall obtain sufficient control over the Digital Objects to preserve them.
Evaluation of the criterion: The criterion is valid but it should be reformulated. The issue of referencing objects on external sites while they are indispensable to the Repository should be addressed. The issue should be assessed depending on the type of external site involved and its long-term future.
Proposal or comment to CCSDS: It should be written that the Repository must have physical control of **each elementary object** in the SIP.

B1.7
Repository provides producer/depositor with appropriate responses at agreed points during the ingest processes.
Evaluation of the criterion: Clear and indispensable.

B1.8
Repository shall have contemporaneous records of actions and administration processes that are relevant to content acquisition.
Evaluation of the criterion: Clear and indispensable.

8.2.2 B2 Ingest: creation of the AIP

B2.1
B2.1 The Repository shall have an associated, printable definition for each AIP or class of AIPs preserved by the repository that is adequate to fit long-term preservation needs. • **B2.1.1 The repository shall be able to identify which definition applies to which AIP.** • **B2.1.2 The repository shall have a definition of each AIP that is adequate for long term preservation, having all the required components, each of which can be maintained over time.**
Evaluation of the criterion: Clear and justified. Document DA4 defines the following rule: *"The Repository shall have formally defined categories of AIPs for which it must assure long term preservation".*
Proposal or comment to CCSDS: sub-criteria B2.1.1 and B2.1.2 appear superfluous.

B2.2
The Repository shall have a description of how AIPs are constructed from SIPs.
Evaluation of the criterion: Clear and justified. Document DA4 includes an identical rule.

B2.3 and B2.3.1
B2.3 The Repository shall incorporate all accepted SIPs into one or more AIPs or otherwise dispose of them in a recorded fashion. • **B2.3.1 The Repository shall follow documented procedures if a SIP is discarded and shall indicate why the SIP was discarded.**
Evaluation of the criterion: Criterion 2.3 appears valid. However, criterion B2.3.1 is to be grouped with B1.5 and reformulated as necessary. If the SIP complies (B1.5), the traceability elements must show how it was reused to build the AIP. If the SIP does not comply, the traceability elements must record the causes of this non-compliance.

B2.4

B2.4 The Repository shall have and uses a convention that generates persistent, unique identifiers for all AIPs.
• **B2.4.1 The repository shall uniquely identify each AIP within the repository.**
  ○ **B2.4.1.1 The repository shall have unique identifiers.**
  ○ **B2.4.1.2 The repository shall assign and maintain persistent identifiers of the AIP and its components so as to be unique within the context of the repository.**
  ○ **B2.4.1.3 Documentation shall describe any processes used for changes to such identifiers.**
  ○ **B2.4.1.4 The repository shall be able to provide a complete list of all such identifiers and do spot checks for duplications.**
  ○ **B2.4.1.5 The system of identifiers shall be adequate to fit the repository's current and foreseeable future requirements such as numbers of objects.**

Evaluation of the criterion: indispensable but some sub-criteria need clarifying: • B2.4.1 only emphasizes the uniqueness of the identifier whereas the sub-criteria address the persistent nature of the identifiers. If this persistence issue is added to B2.4.1, the criterion is identical to B2.4, • B2.4.1 addresses the allocation of a unique AIP identifier, but B2.4.1.2 also mentions identifiers for AIP components. Document DA4 uses the following rule: "*The Repository shall have defined a naming convention by which unique and persistent identifiers are allocated to archived objects".*
Proposal or comment to CCSDS: it is not necessary to have two additional hierarchical levels in this criterion.

B2.4.2
The repository shall have a system of reliable linking/resolution services in order to find the uniquely identified object, regardless of its physical location.
Evaluation of the criterion: This criterion is clear and useful but must not be a second-level criterion.
Proposal or comment to CCSDS: Move this criterion up to the highest hierarchical level. The use of unique and persistent identifiers is a prerequisite to any object resolution service, but this issue is quite distinct from the allocation of identifiers. The resolution system is very important in the area of publications and access to documentary objects of any kind in institutional libraries or repositories. At the same time, there is not the same need for other types of data such as time series for scientific observations.

B2.5
B2.5 The Repository shall have access to necessary tools and resources to provide authoritative Representation Information for all of the digital objects it contains. • **B2.5.1 The repository shall have tools or methods to identify the file type of all submitted Data Objects.** • **B2.5.2 The repository shall have tools or methods to determine what Representation Information is necessary to make each Data Object understandable to the Designated Community.** • **B2.5.3 The repository shall have access to the requisite Representation Information.** • **B2.5.4 The repository shall have tools or methods to ensure that the requisite Representation Information is persistently associated with the relevant Data Objects.**
Evaluation of the criterion: Clear and justified.

B2.6
B2.6 The Repository shall have documented processes for acquiring Preservation Description Information (PDI) for its associated Content Information and acquires PDI in accordance with the documented processes. • **B2.6.1 The repository shall have documented processes for acquiring PDI.** • **B2.6.2 The repository shall execute its documented processes for acquiring PDI.** • **B2.6.3 The repository shall ensure that the PDI is persistently associated with the relevant Content Information.**
Evaluation of the criterion: Clear and indispensable

B2.7
B2.7 The Repository shall ensure that Content Information of the AIPs is understandable for their Designated Community at the time of creation of the AIP. • **B2.7.1 The Repository shall have a documented process for testing understandability for their Designated Communities of the Content Information of the AIPs at their creation.** • **B2.7.2 The repository shall execute the testing process for each class of Content Information of the AIPs.** • **B2.7.3 The repository shall bring the Content Information of the AIP up to the required level of understandability if it fails the understandability testing.**
Evaluation of the criterion: Especially indispensable for complex scientific data. Document DA4 includes the following rule: *"The Repository shall implement a process to check that archived objects are understandable and useful to the designated community".*

B2.8
The Repository shall verify each AIP for completeness and correctness at the point it is created.
Evaluation of the criterion: Clear and justified. An AIP may be incomplete because an expected SIP has not been transferred. This takes us back to the proposal made early in section B1. Document DA4 includes the same rule.

B2.9
The Repository shall provide a mechanism for verifying the integrity of the repository collection/content.
Evaluation of the criterion: Useful and valid with respect to collection. The collection/content combination needs clarifying.
Proposal or comment to CCSDS: “The Repository shall provide a mechanism for verifying the integrity of the repository collections” seems to us more appropriate.
B2.10
The Repository shall have contemporaneous records of actions and administration processes that are relevant to AIP creation.
Evaluation of the criterion: Clear and justified.

8.2.3 B3 Preservation planning

B3.1
The Repository shall have documented preservation strategies relevant to its holdings.
Evaluation of the criterion: Clear and indispensable.

B3.2
B3.2 The Repository shall have mechanisms in place for monitoring its preservation environment. • **B3.2.1The Repository shall have mechanisms in place for monitoring and notification when Representation Information is inadequate for the Designated Community to understand the data holdings.**
Evaluation of the criterion: Clear and justified.

B3.3
B3.3 The Repository shall have mechanisms to change its preservation plans as a result of its monitoring activities. • **B3.3.1 The Repository shall have mechanisms for creating, identifying or gathering any extra Representation Information required.**
Evaluation of the criterion: Clear and justified. The issue of the Repository's ability to manage changes was already addressed in section A. The respective roles of criteria in section A and section B3 should be clarify.
Proposal or comment to CCSDS: Clarification is needed on the position of criteria B3.2 and B3.3 with respect to criterion A3.2.1 (Repository shall have mechanisms for review, update, and ongoing development of its Preservation Policies as the repository grows and as technology and community practice evolve). There is a certain amount of duplication between criteria for fear of forgetting to address certain issues.

B3.4
The Repository shall provide evidence of the effectiveness of its preservation activities.
Evaluation of the criterion: Rather vague and already covered by a host of other criteria To prove that preservation is effective, it may be shown that: 1. All Repository data are accessible, whatever their date of deposit (see B6), 2. The corresponding representation information is understandable by the user community (a point already covered by criterion B3.2)
Proposal or comment to CCSDS: Remove criterion B3.4 or make it more specific to distinguish it from existing criteria.

8.2.4 B4 AIP Preservation

B4.1
B4.1 The Repository shall have specifications for how the AIPs are stored down to the bit level. • **B4.1.1 The Repository shall preserve the Content Information of AIPs.** • **B4.1.2 The Repository shall actively monitor the integrity of AIPs.**
Evaluation of the criterion: Valid, but the criteria hierarchy is not coherent. The top-level criterion must cover all the subjects defined below it. On top level B4.1), only the specification issue is addressed. At lower levels (B4.1.1 and B4.1.2), the implementation of preservation procedures is addressed. In this regard, document DA4 has defined only one rule about preserving bits on a medium: *"The Repository shall have defined a documented strategy for data archival and migration of the medium used to store archived data*".
Proposal or comment to CCSDS: Broaden the scope of B4.1 to take into account implementation of preservation. *Furthermore, it is necessary to clarify the role of this criterion with respect to A3.5:* ***Repository commits to defining, collecting, tracking, and appropriately providing its information integrity measurements.***

B4.2
B4.2 The Repository shall have contemporaneous records of actions and administration processes that are relevant to storage and preservation of the AIPs. • **B4.2.1 The Repository shall have specifications for all actions taken on AIPs,** • **B4.2.2 The Repository shall be able to demonstrate that any actions taken on AIPs were compliant with the specification of those actions.**
Evaluation of the criterion: Valid and justified.

8.2.5 B5 Information management

B5.1
The Repository specifies minimum Descriptive Information requirements to enable the designated community to discover and identify material of interest.
Evaluation of the criterion: Clear and indispensable. Document DA4 contains the following rule: *"The Repository shall have identified and collected metadata enabling the user community to find useful data, assess their suitability for user needs then use them".* *It also adds a rule on the implementation of search procedures:* *"The Repository shall have defined and implemented processes to search for and extract the content of metadata geared to user needs and its own needs".*

B5.2
The Repository captures or creates minimum Descriptive Information and ensures that it is associated with the AIP.
Evaluation of the criterion: Clear and indispensable.
Proposal or comment to CCSDS: Criteria B5.1 and B5.2 could be regrouped in the same way as B4 .1 covers B4 .1.1 and B4.1.2

B5.3 et B5.3.1
B5.3 The Repository shall maintain bi-directional linkage between each AIP and its descriptive information. • **B5.3.1 The Repository shall maintain the associations between its AIPs and their descriptive information.**
Evaluation of the criterion: B5.3 is clear and indispensable. B5.3.1 is superfluous because B5.3 already contains the verb "maintain"
Proposal or comment to CCSDS: remove criterion B5.3.1

8.2.6 B6 Access management

B6.1
B6.1 The Repository shall comply with Access Policies • **B6.1.1 The Repository shall log all access management failures, and staff review inappropriate incidents.**
Evaluation of the criterion: Clear and indispensable for B6.1. The logging of access failures seems to us to be part of general supervision of the technical infrastructure to ensure correct operation. Document DA4 is far more specific on the question of access, with the following three rules: *"The Repository shall make available to the user community information on the various search and retrieval possibilities for archived data, and the corresponding conditions in terms of rights, rules of use and costs".* *"The Repository shall implement logical protection and authentication mechanisms that comply with the restrictions and access conditions defined by the Project".* *"The Repository shall found its mechanisms for data searches on open query, search and retrieval standards".*
Proposal or comment to CCSDS: There is some inhomogeneity among criteria. Some are very precise, like the criterion on unique and persistent identifiers, whereas others are vague, like B6.1 on access.

B6.2
B6.2 The Repository shall follow policies that enable the dissemination of copies of that are traceable to the originals, with evidence supporting their authenticity. • **B6.2.1 The Repository shall record and act upon problem reports about errors in data or responses from users.**
Evaluation of the criterion: Clear and justified. The need to be able to restore an authentic copy must, however, be specified in the Repository's mission statement or in the deposit agreement with the producers. It is particularly important for issues relating to intellectual property and any digital object with evidential value.

8.3 SECTION C - INFRASTRUCTURE AND SECURITY RISK MANAGEMENT

A non-negligible part of the criteria is covered by CNES's ISO 9001 certification and the constant quality enhancement process in which CNES is engaged.

We can quote here a set of processes currently defined and applied in the CNES management system:

For information management:

- Process S3 on control of the CNES information system, which includes aspects such as guaranteeing and maintaining information contents,
- Organization of documentation at CNES: CNES-SQ-NO-145 of June 2005,
- The guide on setting up a project's technical documentation management system, DCT-SQ-NO-322 of December 2007,
- CNES information assets management, CNES-SQ-NO-501 of May 2008,

For security:

- Process S6 on protection and security of information systems, description of the process, process management data sheet,
- General decision CNES/P/2007-51 on the organization of information system security,
- Procedure "Establishing and checking project security requirements", CNES-SQ-PR-207 of January 2007,
- Procedure "Implementation of protection resources", CNES-SQ-PR-208 of January 2005,
- The security approval documents of Sipad-NG,
- Documents on controlling physical access to the site, premises, infrastructures and corresponding protection means,
- Fire protection,

For the infrastructure used by the Repository:

- Process S4 on life-cycle support for infrastructures and resources: this process includes planning activities, supplies and services, preventive and corrective maintenance, upgrading, supervising repository operation etc.

8.3.1 C1 Technical infrastructure risk management

C1.1
The Repository shall identify and manage the risks to its preservation operations and goals associated with system infrastructure.
Evaluation of the criterion: Clear and indispensable. However, this criterion on risk management in preservation operations must be clearly positioned with respect to C2.1, risk management for information security. See below.
Proposal or comment to CCSDS: Risk management is a general task. Splitting this section into two parts - one on risks relating to software, hardware, storage media and operational procedures (C1.1) and the other on risks relating to information security (C2.1: the Repository maintains a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant.) - should be justified because the boundary between the two areas appears nebulous. Risk management methodologies always involve analysing the context, identifying all risks, assessing the probability and impact of these risks then taking a decision and implementing the chosen actions. *The only apparent justification for such a separation is the existence of requirements and standards on good practices concerning security risks.* *This remark targets the structure and hierarchy of criteria, not their content.*

C1.1.1 (hardware and operating systems)
C1.1.1 **The Repository shall perform technology watch.** C1.1.1.1 **The Repository shall have hardware technologies appropriate to the services it provides to its designated communities.** C1.1.1.2 **The Repository shall have procedures in place to monitor and receive notifications when hardware technology changes are needed.** C1.1.1.3 **The Repository shall have procedures in place to evaluate when changes are needed to current hardware.** C1.1.1.4 **The Repository shall have procedures to replace hardware when evaluation indicates the need to do so.**
Evaluation of the criterion: Clear and justified.
Proposal or comment to CCSDS: It appears useful to separate hardware from software. The technology watch and change management procedures are different and are handled by distinct teams.

C1.1.1 (software)
C1.1.1 **The Repository shall perform technology watch.** C1.1.1.5 **The Repository shall have software technologies appropriate to the services it provides to its designated communities.** C1.1.1.6 **The Repository shall have procedures in place to monitor and receive notifications when software changes are needed.** C1.1.1.7 **The Repository shall have procedures in place to evaluate when changes are needed to current software.** C1.1.1.8 **The Repository shall have procedures to replace software when evaluation indicates the need to do so.**
Evaluation of the criterion: Clear and justified.
Proposal or comment to CCSDS: see the comment on the previous criterion.

C1.1.2
The Repository shall have adequate hardware and software support for backup functionality sufficient for preserving the repository content and tracking repository functions.
Evaluation of the criterion: Clear and indispensable. These aspects are covered by process S4 of ISO 9001 certification (life-cycle support for infrastructures and resources).

C1.1.3
C1.1.3 The Repository shall have effective mechanisms to detect bit corruption or loss. **C1.1.3.1 The Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.**
Evaluation of the criterion: Clear and indispensable.

C1.1.4
The Repository shall have a process to react to the availability of new security updates based on a risk-benefit assessment.
Evaluation of the criterion: It is not clear to us why this aspect is not covered by C2, criteria on security risks.
Proposal or comment to CCSDS: re-examine the distribution of criteria relating to security.

C1.1.5
The Repository shall have defined processes for storage media and/or hardware change (e.g., refreshing, migration).
Evaluation of the criterion: Clear and indispensable.

C1.1.6
C1.1.6 The Repository shall have identified and documented critical processes that affect its ability to comply with its mandatory responsibilities. **C1.1.6.1 The Repository shall have a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities** **C1.1.6.1.1 The Repository shall have a process for testing the effect of changes to the repository's critical processes.**
Evaluation of the criterion: Clear and indispensable.

C1.2 et C1.2.1
C1.2 The Repository manages the number and location of copies of all digital objects. **C1.2.1 The Repository shall have mechanisms in place to ensure any/multiple copies of digital objects are synchronized.**
Evaluation of the criterion: Clear and indispensable.

8.3.2 C2 Security risk management

C2.1 and C2.2
C2.1 The Repository maintains a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant. **C2.2 The Repository shall have implemented controls to adequately address each of the defined security risks.**
Evaluation of the criterion: Clear and indispensable.

C2.3
The Repository staff shall have delineated roles, responsibilities, and authorizations related to implementing changes within the system
Evaluation of the criterion: Clear and indispensable.

C2.4
The Repository shall have suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an offsite copy of the recovery plan(s).
Evaluation of the criterion: Clear and indispensable. It should specify that the off-site backup should not be in the same geographical area.

Complementary elements:

In the document applicable to CNES digital repositories, we defined the following rule:

"The data search and retrieval functions shall not have any sizeable impact on user hardware or network resources".

Justification: The Repository must gear its offer to user community needs. It would be inacceptable to place any constraints on user hardware, and could be totally incompatible with the constraints of the institution to which the user belongs. The Repository should be able to tailor its data distribution resources to users' network constraints. In practical terms, this means that the user should not be obliged to use a particular navigator or install a specific software program.