

**SERAD (CNES Service for Data Referencing and Archiving) Main Author Full  
Danièle Boucon(1), Richard Moreno(1), Dominique Heulet(1), Paul Kopp(1), Michel Duplaa(1),  
Martine Larroque(1)**

<sup>(1)</sup> *CNES*

*18 av E. Belin, 31401 Toulouse Cedex 9, France*

*EMail: Daniele.Boucon@cnes.fr*

## **ABSTRACT**

SERAD (Service for Data Referencing and Archiving) is a project that will allow CNES to dispose of a centralized repository structure that collects, stores and disseminates information and metadata on Data that are under CNES responsibility. The objective for CNES is to better handle and to improve the access to this data patrimony.

Data, referencing, archiving, clearinghouse, access, oais

## **INTRODUCTION**

SERAD (Service for Data Referencing and Archiving) is a project that will allow CNES to dispose of a centralized repository structure that collects, stores and disseminates information and metadata on Data that are under CNES responsibility. The objective for CNES is to better handle and to improve the access to this data patrimony.

In order to achieve this goal, it is mandatory to identify all data which are relevant and to verify whether these data are properly archived ; if not, then, to proceed to the archiving of these data.

The SERAD mission is then the following :

- To constitute and maintain an open and centralized metadata repository of all data that are under CNES responsibility,
- To do, when necessary, the archiving of data
- To survey the data production centres in order to guaranty the long term preservation of these data even if (critical case) one of these centres has to be closed.

This system will be built upon existing generic tools that will be customized :

- BDMS which is a kind of clearing house not limited to geographic information, but able to handle data from any thematic.
- SIPAD-NG which is a generic tool allowing to give a full access (with search criteria) to (nearly) any kind of archived data.

This paper will present in detail all the objectives of SERAD, the technical architecture and all the work that has already been done.

# 1- PRESENTATION OF SERAD PROJECT

Some years ago, CNES engaged a reflection about long term preservation of scientific data. For example, this resulted by the creation of Thematic Poles as ICARE, ETHER, CDPP, ... and more specifically in the field of plasma physic (CDPP). This reflection is now extended to all data which are under CNES responsibility, with the following problematic :

patrimonial

how to inform the users about the localisation and the usage of data

## 1.1. Background

For 40 years, in CNES, a big number of space missions have been producing a huge amount of data (hundred of To). This data constitutes a valuable heritage that must be preserved because many of them are :

- unique when related to an event that will never happen again or in a very long time (e.g. Halley comet period is 76 years !)
- integrated in long cycles of observations, including cycles climate change observation;
- mandatory to prepare for future missions (e.g. GAIA take benefits for HIPPARCOS experience)

With the arrival of new missions, this amount of data will further increase in volume and complexity. This is mainly due to:

- geographical dispersion of data: the missions are often conducted with partners and the data produced are not necessarily stored at CNES; when they are, they are not always in the same place;
- variety of thematic (science of the universe, earth science, life science, microgravity, etc.).
- variety, at a given time, of data production and of data preservation techniques "state of the art"

## 1.2. Interest of the Service Referencing and Archiving Data (SERAD)

The CNES SERAD offers:

- to add values to its data patrimony allowing that it can be used;
- to promote the data patrimony up to the widest public, wider than only people that contributed to the creation of this data, thus participating in CNES promotion;
- effective control, under its two functions of agency program and technical center of all its data (inventory, location, description, ...);
- a reduction in the long term of data preservation costs through the control of all data and via a dedicated organization

## 1.3. SERAD Mission

The problem of the researcher, and more generally any user data and archived documents, is to recover the data directly relevant to his need. Therefore he shall be able to query the data patrimony and receive an appropriate response, not tainted by error or omission to his query.

It is well known that a poorly preserved data, or a well preserved but poorly described data, will lead to a loss of the usability of the data becomes.

The SERAD is a service that ensures the integrity of heritage data from CNES. For this it

1. constitutes and maintains a reference of data that are under CNES responsibility (it allows users to find data relevant to their needs), it informs users of what exists and how to access it;
2. archives data, whenever long term preservation is not (or not anymore) supported by a data center that would be responsible. It thus retains data without degradation in the long term, together with all information necessary for intelligibility and their employment;
3. provides a monitoring service vis-à-vis the data production centers of data to prepare and organize their data long term preservation activities

## 2- SYSTEM DEFINITION

### 2.1. General definition

The overall mechanism of referencing and archiving CNES data from CNES to put in place in order to ensure the SERAD is based on building blocks:

- the BDMS (Bureau des Métadonnées et des Services), 1. The BDMS is a referencing tool (clearing house) installed on the CNES Information System Direction machines. It will allow to describe accurately the data sets and services that can be associated to them,
- the SIPAD-NG (Système d'Information, de Préservation et d'Accès aux Données Nouvelle Génération). This is a generic tool for data management, which can operate in different contexts. It implements the full OAIS model.

The SERAD interfaces with data centers in different themes, internal or external to CNES.

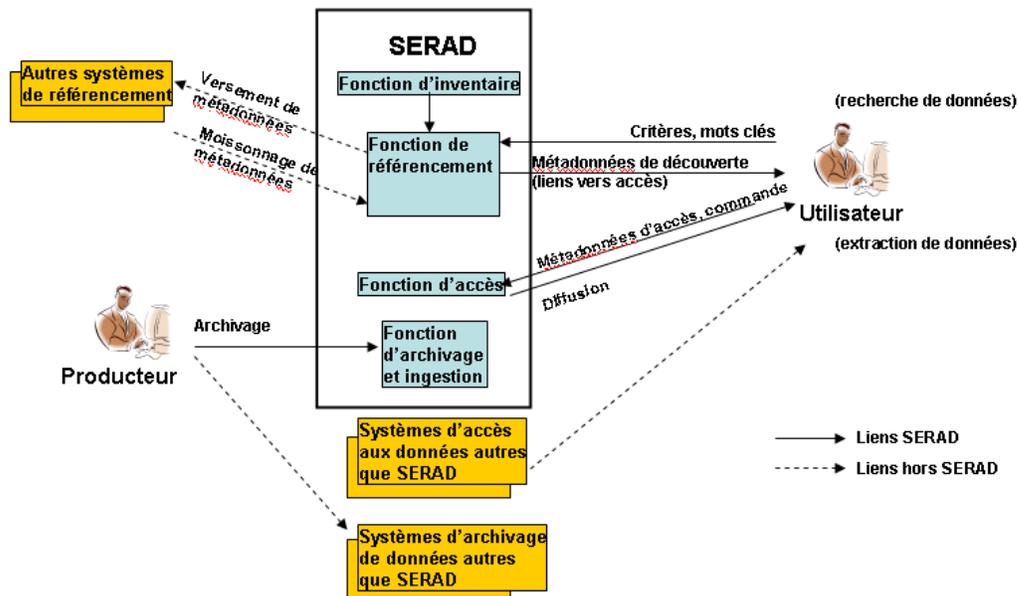
The SERAD uses the common means of CNES Information Systems Direction (computing resources, storage facilities), including the STAF (Service de Transfert et d'Archivage de Fichiers) for archiving files.

As stated above, the SERAD is not a fully developed system, its implementation is essentially evolutions and adaptations of existing building blocks.

Future work will also be based on existing standards or norms (OAIS, ISO 19115, ARK, ...).

### 2.2. Functional definition

The system consists of 4 main functions :

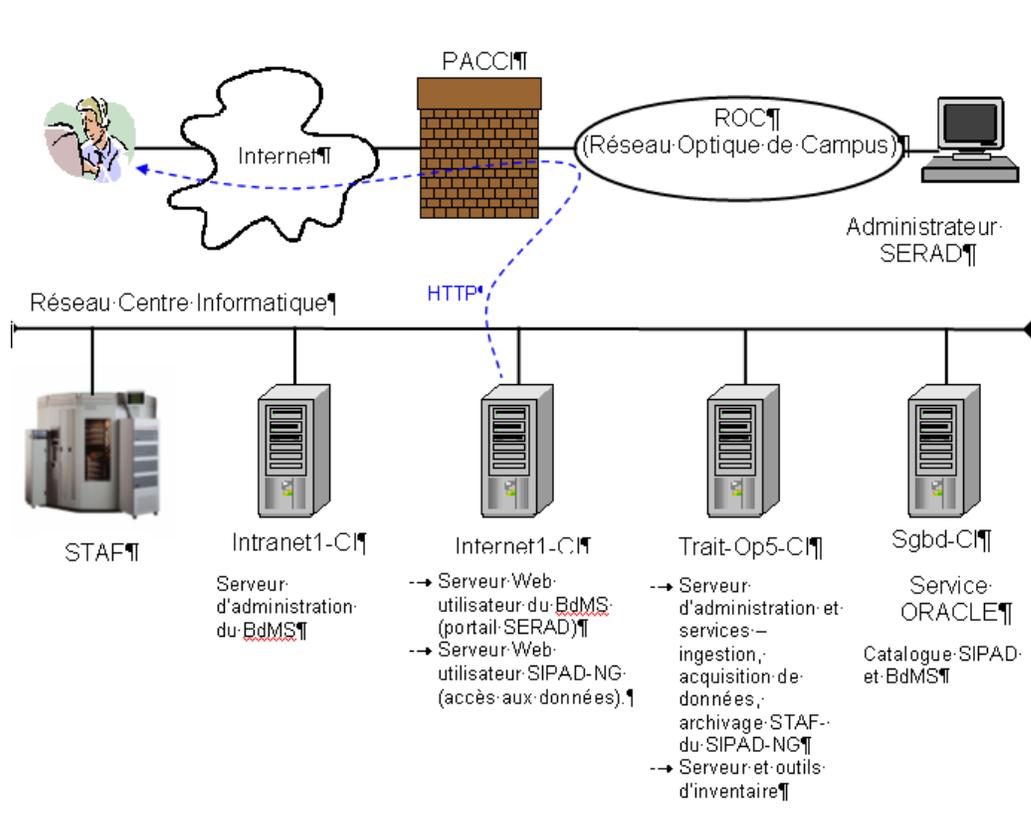


- The referencing function. She allows the user to search data of interest through the discovery metadata, using criteria (interests) and keywords. The function of referencing returns to the user the metadata for its selection. These metadata contain descriptive information of potentially interesting data,

and the link to the data access function. This function will also broadcast, harvest other systems of referencing, as the IDN for example.

- The access function: it allows the user to order data. Its control is done using metadata access (on temporal criteria for example). Other services can be provided at that time (processing format for example). The access function allows the controlled release of data to the user.
- The archiving function and ingestion: it allows archiving data from a producer and ingestion of metadata access in the SIPAD-NG catalogue.
- The inventory function: it provides an exhaustive inventory of the candidate data for referencing.

### 2.3. Architecture

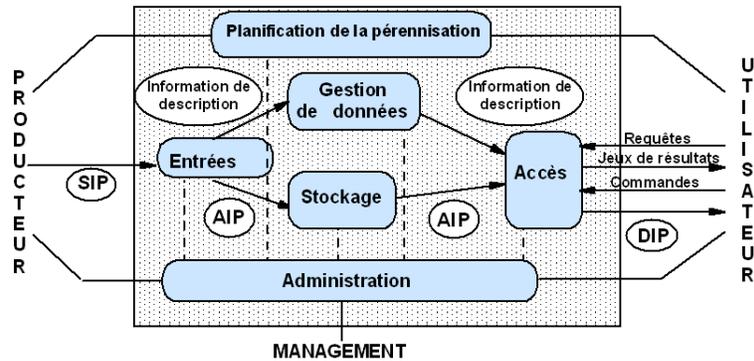


The SERAD technical solution is based on the CNES Information System Direction infrastructure. There will be

- One SIPAD-NG instance on a computer accessible through Internet
- One BdMS instance, on the same computer
- The STAF, which is a big redounded tape library,
- The inventory server that will be installed on an Intranet computer

## 2.4. Tools

- SIPAD-NG is a generic tool for data archive management. It implements the full OAIS model.



SIP = Submission Information Package / *Paquet d'information à verser*  
 AIP = Archival Information Package / *Paquet d'informations archivé*  
 DIP = Dissemination Information Package / *Paquet d'informations diffusé*

This tool is currently used in SSALTO (altimetric products), CDPP (plasma physics thematic pole), MERCATOR (ocean products), DECLIC (microgravity experience) and SMOS (Soil Moisture and Ocean Salinity).

The screenshot shows the AVISO CNES Data Center website. The header includes the AVISO logo and a satellite image. A navigation menu on the left lists 'DATA SELECTION', 'SERVICES', 'USER', and 'HELP'. The main content area features a banner stating 'This server gives access to CNES altimetry products' and lists various satellite missions: TOPEX/Poseidon, Jason, SPOT5, ERS, OSTIM/JASON2, and ENVISAT. A registration warning is present, and the footer includes logos for SALP, SAO, SIPAD, CNES, and CLS.

- BDMS is a tool which aims at allowing users to discover that a datasets of its center of interest exists.

This tools is based on the ISO19115 standards.

This standards is dedicated to Geographic Information. It will then be profiled and enriched in order to be able to fit the needs of all thematic. This metadata model enrichment will be based on the thematic classification and interview work described in section 3 of this paper.

**Request** (Feed this form to submit a request)

**Selected profile** **Profil1**

Proposition	Selected
CRIL	CRIL
CNES	CNES

Clearinghouses to search:

**Ordinal position**

90.0 N  
180.0 W 180.0 E  
90.0 S

Keyword	discipline	place	stratum	temporal	theme
Thesaurus1	Key2 Key3	Key6	Key3 Key4	Key8	Key1 Key2
Thesaurus2		Key9			Key1

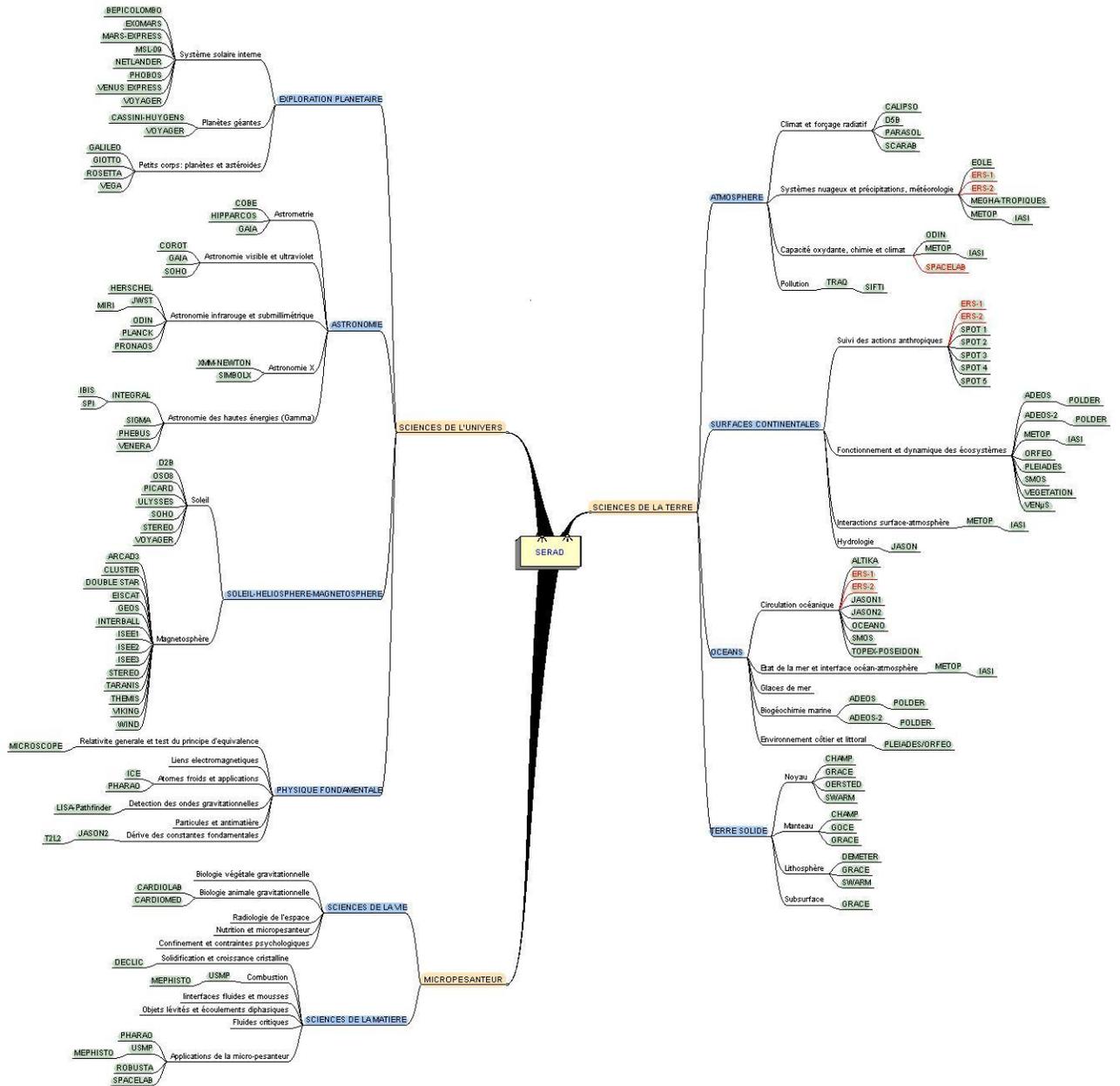
Period	Season period
Start date: YYYY/MM/DD	Start date: DD/MM
End date: YYYY/MM/DD	End date: DD/MM

Feature Type	Full Search	Text
<input type="text"/>	<input type="text"/>	<input type="text"/>

### 3- DEFINITION OF A MINIMAL METADATA MODEL

This work includes the definition of:

#### 3.1. Classification



The objective is to define the centers of interest which are representative, discriminating and consistent to the terminology of the future users. This required a precise classification job of all the science thematics. This has been done thanks to the contribution of all CNES thematic responsible from the CNES Programmes Direction. This work is also done with the help of the C-S company

## **3.2. Thesaurus**

The goal is to have an ordered list of keywords recognized by communities of users in order:

- to speak a common language understandable to those communities,
- to be interoperable with other systems, precisely target datasets.

This work is conducted in parallel with the classification of science thematics. It is a difficult exercise which will be expanded over time, data sets that will be introduced gradually in the SERAD, and from elements collected so far (thesaurus CNES, ...). This work is conducted with the support of CS-SI.

3. Exhaustive list of duties and assignment of classification

## **3.3. Exhaustive list of missions**

The goal is to have an exhaustive list of CNES missions, and to be able to decide what to do with the datasets. These missions are also integrated in the classification tree.