SERAD (CNES Service for Data Referencing and Archiving)

Martine Larroque[1]Danièle Boucon[1], Richard Moreno[1],

Dominique Heulet[1],

[1] *CNES*
18 av E. Belin, 31401 Toulouse Cedex 9, France
EMail: martine.larroquen@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

1 - BACKGROUND

Some years ago, CNES engaged a reflection about long term preservation of scientific data. 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

Why is so important to preserve the sustainability of scientific data, the most important points are :
- The complexity of instruments and space systems
- The objectives of scientific communities in space
- The term science projects
- The quantity and quality of information produced
- Not reproducible data (events: comets, OT)
- The cost of space projects
- Broad justification for the scientific and technical value of spatial data
- Observation: the data is located across themes and multi.

2 –THE COMMUNITIES OF USERS

We can identify several categories of users or communities of users.

- CNES/CNRS/INSU scientific community in order to :

    - Preparation of future missions
    - Access to Heritage spatial data
    - Heritage enhancement of spatial data (application services - development, studies, statistics, education, ...)

- Explication of changing needs of scientific communities (comparative reprocessing intersection of   multi-thematic, ...)

- Exposure data for other space agencies (ESA, ..) and research organizations

- Access to spatial data (scientists, students, groups, ..)

3 – PRESENTATION OF SERAD PROJECT

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 the missions which are under CNES responsibility and verify whether these data are properly archived ; if not, then, to proceed to the archiving of these data.

## 3.1 The objectives of data preservation in CNES are the following

- Keep in time, information from instruments and space systems sometimes complex

- Meeting the research objectives and use of spatial data for scientific communities over the long term(largetime series, episodic events or non-repeatable)

- Ensuring long-term quantity and quality of data produced during missions

- Keep information accessible and understandable to sustainably in order to extract the maximum value of scientific or technical

- Ensuring sustainable access for reuse without loss of data from scientific missions

- Reduce long term costs of preservation through control of all data and a dedicated organization

3.2. The Service Referencing and Archiving Data (SERAD)

The SERAD missions are the following:

- Constitute and maintain a centralized repository of all data that are under CNES responsibility,

- 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;
   Make long term archive, if necessary,

-    Make surveys in the data production centers in order to prepare and guaranty the long term preservation of these data in particular in the case of end or stop of data production in the centers

3 – DATA MANAGEMENT
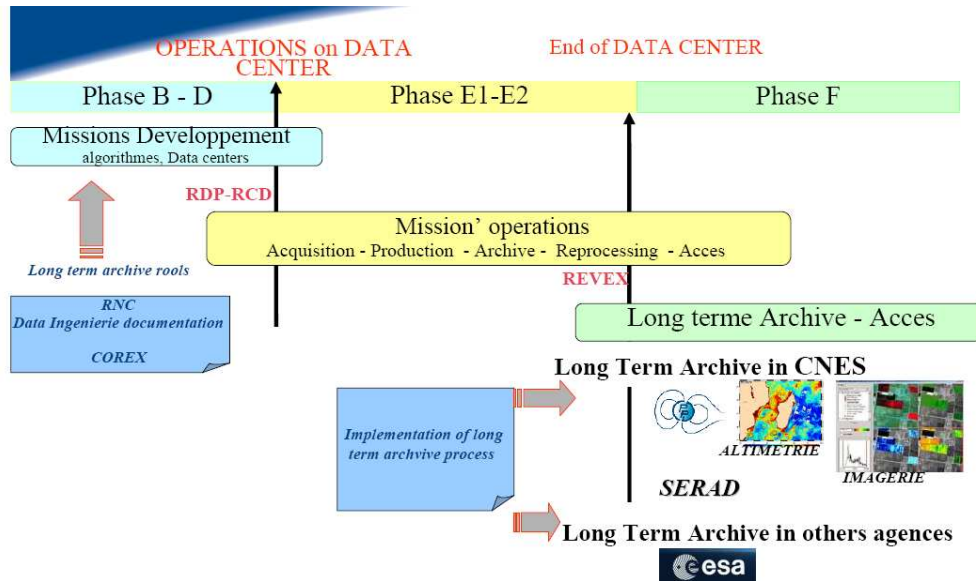
3.1 The objectives of data management

The objectives of data management need a defined data policy.

> - Clear strategy in terms of data policy, in particular, what data should be made permanent, (CNES legacy data, laboratory data, data generated with CNES participation, etc.)
>
> - Meet the needs of the scientific community (SPC, TOSCA, CERES)
>
> - Ensure consistency with the many groups international standards and standardization (CCSDS, ISO, LTDP ...)
>
> - Answer questions about the functions for archiving and sustainability

- What scope of missions should be taken into account?
- How to set priorities in the missions to be taken into account?
- How much data should we sustain?
- What criteria do we decide to archive data sets?
- How to ensure their use?

-→ Mission's categorisation
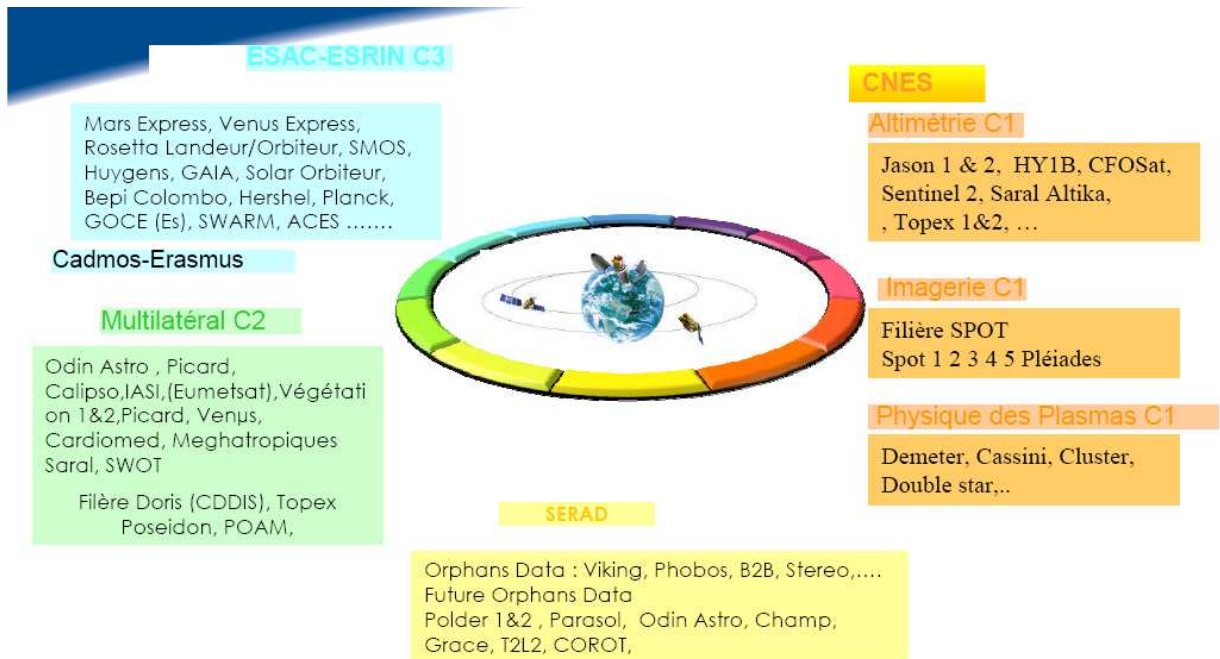
3.2 - Long term data preservation process

The following parts were identified in the process:

A- Characterization of the mission statement in order to deduce the involvement of CNES implementation of the policy data

B – Preliminary Study of sustainability of mission data
Elements to be preserved: N0, N1, N2, processing lines (algos), auxiliary data, volume, data sets, data formats, documentation, contacts managers scientists, ... ◊ Inventory Sheet Scientific or historic → Options and cost assessment of sustainability

C - Decision on continuation of the preservation of data in a Steering Committee

D - Implementation of the sustainability process data
    Recovery and data preparation, data archiving, metadata creation, data dictionary, criteria for data access, data access (public / private) format conversion - Plan Writer Archive

Categorisation -→ Preliminary Studies -→ Steering Committee→ Implementation
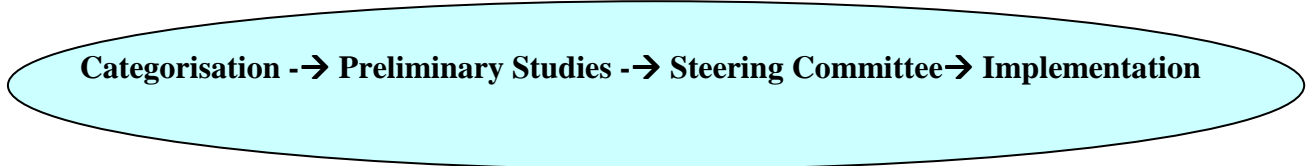
3.2 – Missions Pilots

In a preliminary work, about 180 missions with CNES's participation have been inventoried in 3 thematics Earth Observation, Sciences of the Universe, Microgravity.

In order to develop the data management and qualify tools, pilot missions have been chose.

One by thematics :

- Polder for Earth Observation,
- Odin for Sciences of the Universe
- Mephisto for Microgravity.
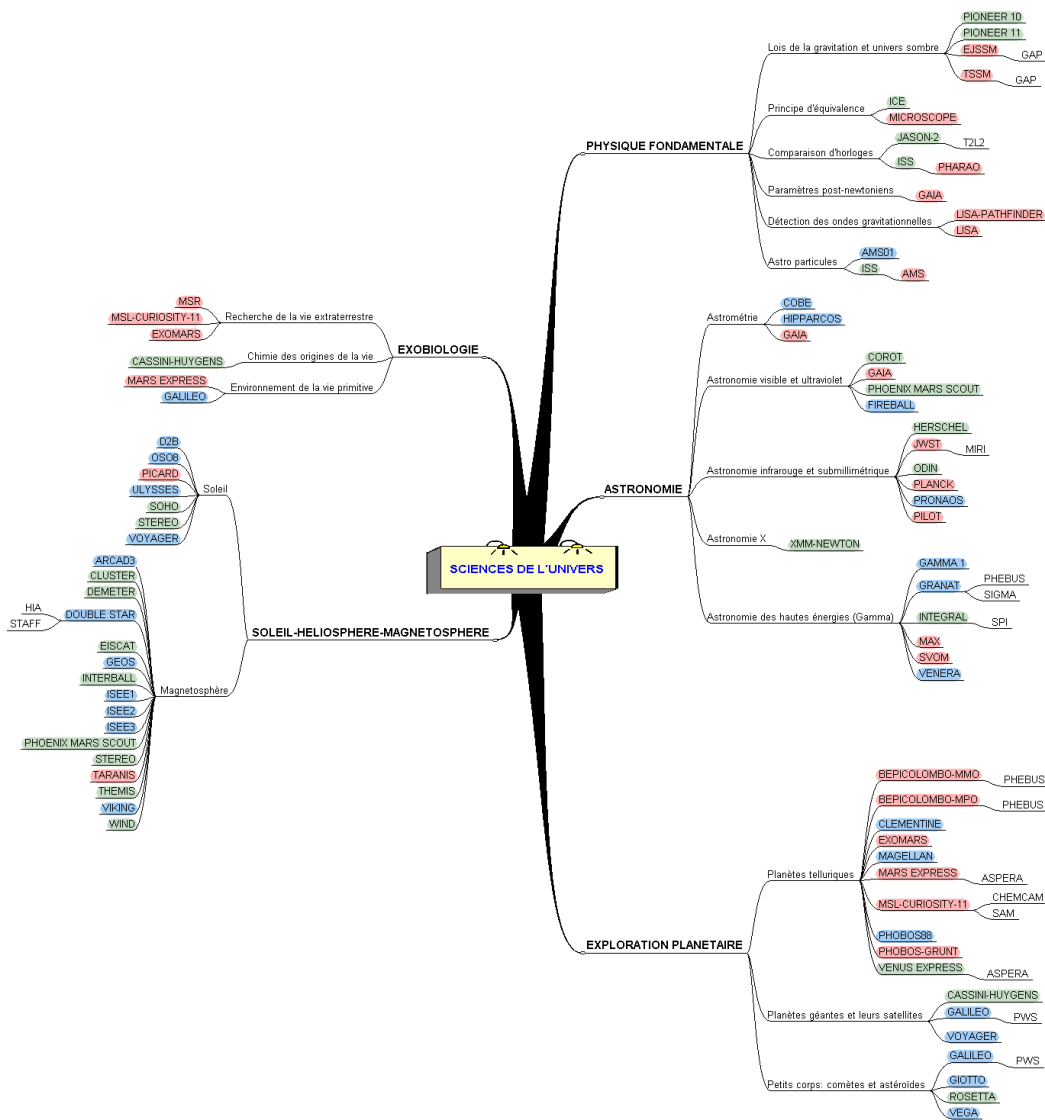
3 – OVERVIEW OF TOOLS USED IN THE PROJET

The overall mechanism of referencing and archiving CNES data from CNES to put in place in order to ensure the SERAD is based on basic components .

• The **SERAD Clearing house** is a referencing tool (clearing house) installed on the CNES Information System Direction machines. It will allows to describe accurately the data sets and services that can be associated to them, ISO 19115 Geographic Information Metadata (ISO), XML (inventory sheet).

• 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 **BEST and Scribe Tools** used to describe the data formats and produce Data ICD if necessary.

 - The **Dash Board** used to plan the missions in order to be taken into account by the archive.
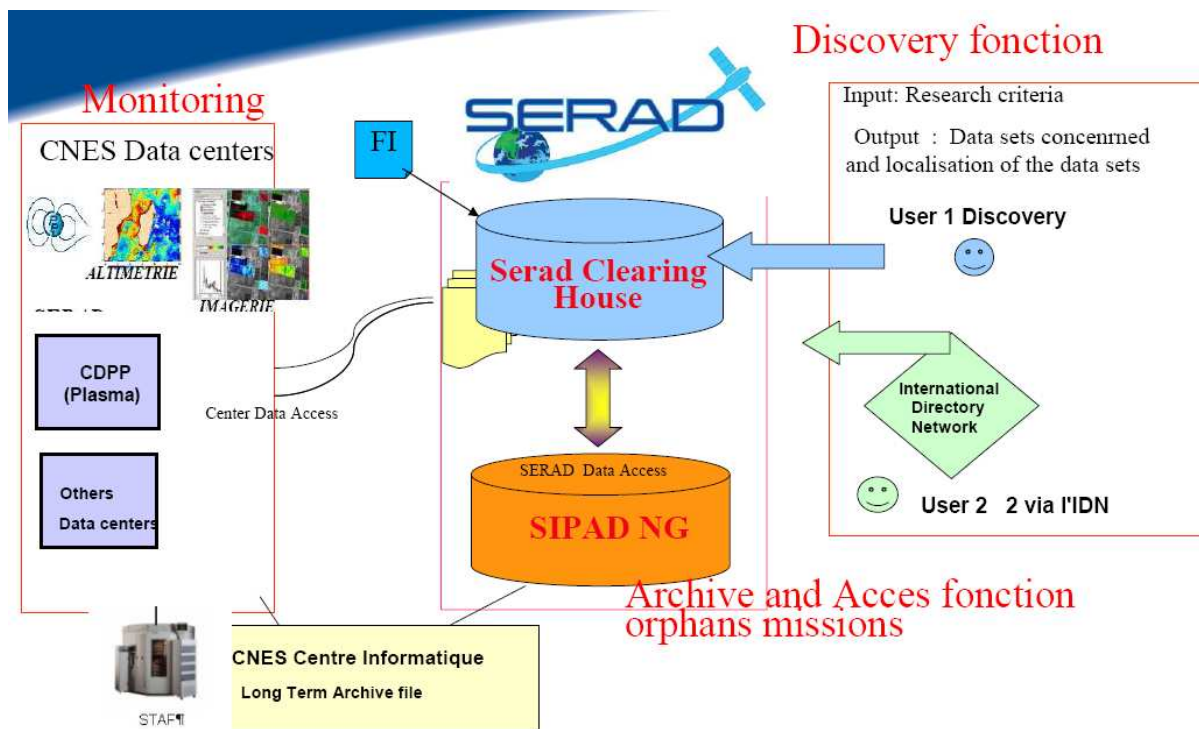
The SERAD interfaces with data centers in different themes, internal or external to CNES.

The SERAD uses the common means of CNES Information Systems Direction (computing resources, storage facilities), including the STAF for archiving files.

The SERAD is not a fully developed system, its implementation is essentially evolutions and adaptations of existing basic components.

Activities are also based on existing standards or norms (OAIS, ISO 19115, ARK, ...).
ISO 19115 Geographic Information Metadata (ISO)
XML (inventory sheet)
OAIS for SIPAD NG

SIPAD NG for Aviso Data Center