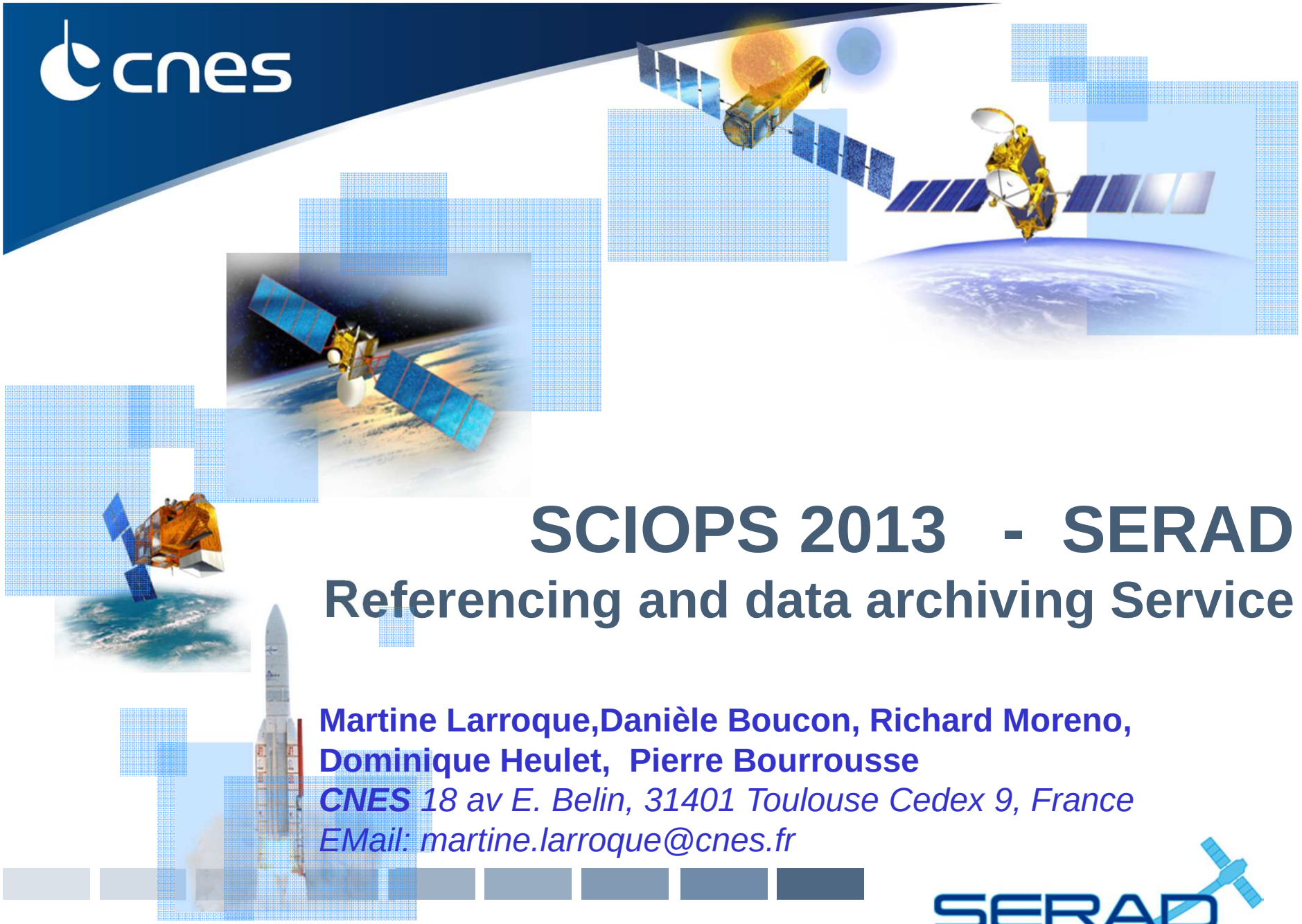


SCIOPS 2013 - SERAD

Referencing and data archiving Service

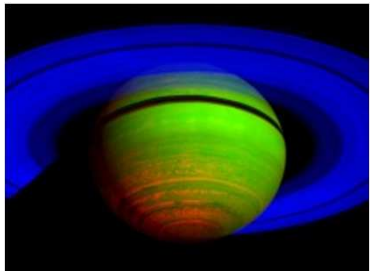
Martine Larroque,Danièle Boucon, Richard Moreno, Dominique Heulet, Pierre Bourrousse
CNES *18 av E. Belin, 31401 Toulouse Cedex 9, France*
EMail: martine.larroque@cnes.fr

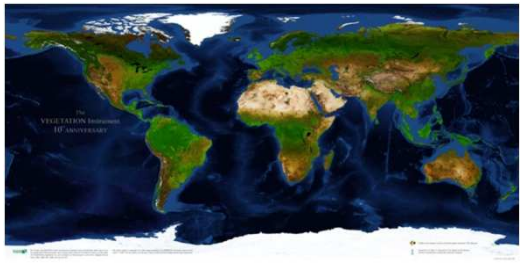
OUTLINE

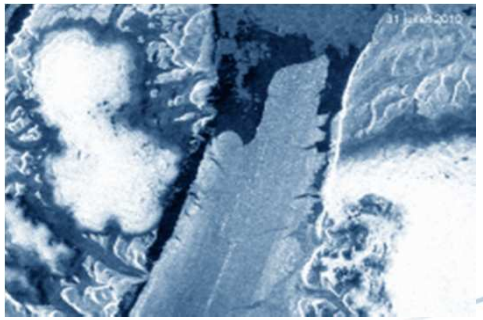
- Objectives of data preservation
- CNES Status on data preservation
- SERAD general presentation
- System SERAD and tools
- Synthesis

LONG TERM DATA PRESERVATION OBJECTIVES

- **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 (long time 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

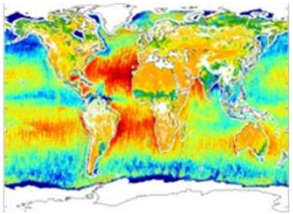






LONG TERM PRESERVATION OBJECTIVES

Scientists



DATA PRESERVATION NEEDS

Scientific communities, UE Directives,…

Management

Lawyers

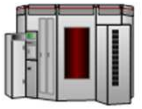
GOUVERNANCE
DATA POLICY
FUNDS

Space Agencies, Laboratories, Partners

Technical teams

LEGACY
NORMATIVES ASPECTS GUIDELINES

PAIMAS, OAIS, ISO



TOOLS, DATA PROCESSING, INFRASTRUCTURE

CNES MISSIONS CURRENT STATUS

Universe Science missions

- Astronomy **Corot** , Solar **Picard**,
- Fundamental physics **T2L2,** Magnetosphere **Demeter,…**
- Plasma : **Eiscat, ISEE3/Ice, Arcad3**,
- Rosetta, **MSL/CHEMCAM & SAM** …. And also …**old missions**

Earth Observation missions

- Multi-spectral imaging missions/sensors, high and very high resolution:
  SPOT 1, 2, 3, 4, 5, Pléiades
- Medium resolution Land & Ocean monitoring missions/sensors:
  JASON 1/2/3, Topex, Altika/Saral
- Atmospheric chemistry missions:
  Polder, Parasol, Calipso, Megathropique,…
- Geodesy topics : **Champ,** …;

and **Balloons experiences,…**

CNES APPROACH ON DATA ARCHIVE

Responsibilities/mandate on data preservation

- For scientific missions: frequently **a convention** exists between CNES and the partners to preserve raw and/or level 1 data.
- Upper levels can also be preserved if necessary
- For **old missions** and if necessary **the legal aspects** must be instructed in the data preservation preparation phase

Time limitation for data preservation

- There is no a time limitation; CNES Mandate

CNES APPROACH ON DATA PRESERVATION

A CNES **data preservation guidelines defined,** centralize and accessible in the "RNC" Référentiel Normatif du CNES" (enterprise repository). How and what must be preserve . Best practices on LTA for projects are in development phases

Our strategy is to **develop generic products**, able to cope with all the sets of thematics

A possible extension of the **infrastructure of Storage STAF** (Example: Pléiades) if necessary

Archive security : the tape library has a back-up which is hosted in another building. Both libraries a located in restricted access areas.

Since 2008, **Data Exploitation is handled by a sub-direction in CNES :**

"Missions Operations" : the organization is made according to 3 themes Earth Observation, Science of Universe, Microgravity

- process data according to users needs,
- perform archive operations,
- Ensure **Referencing & Long Term data Preservation** (for all thematics) :

SERAD project : development phase 2009-2012, operational phase 2013

SERAD OBJECTIVES
Referencing and data archiving Service

Global data referencing with missions on CNES participation (Clearing House)

SERAD will provide for each project managers mission for all topics EO, UE, and Microgravity, at end of life of the mission :

- **Survey and expertise** for data preservation for projects at the end of the production period, after data reprocessing tasks

- **A permanent repository** (when data processing centers have stopped their activities)

- **For a fixed period**, a technical description and financial

The preservation instruction phase begins 2 years before the date of end of mission

2 types : - Orphan missions data (old missions) , no access possible/ archived in CNES

- Data from stop data processing centers (COROT,…)

SERAD OBJECTIVES

REFERENCING ACTIVITIES

- Establish and maintain a centralized repository of missions involvement CNES
- Provide access to the scientific community and CNES

DATA PRESERVATION ACTIVITIES

- Existing orphaned data (25 old projects): absorb the passive of orphan data
- Treatment centers at the end of life (projects whose mission ends: Corot, Polder / Parasol, about 20 projects by the end of 2017 ...)
- Partnering in the case of taking into account the long-term preservation by third parties (agencies, ...)
- Data N0, N1, N2, auxiliary data, software, file description, documentation, ...

- SURVEY PROJECTS & DECISION

- Ensure monitoring role & expertise with operating projects
- Implement the decision process of sustainability

SERAD Archive Management

Before the end of the operational phases (at least 2 years) project request a study on the technical and financial feasibility of sustainability of their data.

An annual session evaluates (by critéria) the potential missions to preserve data and establishes a ranking priority of these missions (orphans mission data and recent mission data)

A steering committee of scientific data ensures the management of the archive, it validates the workload forecast, the resources and planning corresponding and ensures the availability of IT resources

For old MISSIONS → low availablity of scientifics teams to give informations about the missions and data description : ODIN, Phobos, Phebus, OSO8, Hipparcos,……

LONG TERM DATA PRESERVATION PROCESS
Orphans data – End of missions (data centers)

A- Characterization of the mission statement in order to deduce the involvement of CNES implementation of thedata policy - 3 Categories depending on the mode of intervention on the production data: internal CNES, partnerships, agencies, …

B - Preliminary Study of sustainability of mission data

Elements to be preserved: L0, L1, Software (data processing: algorithmic part) Other Data levels may be archived depending on the project: L2, L3, auxiliary and anciliary data, data volume, data sets, data format (FITS, CLASS), documentation, scientists contacts for knowledge aspects , …) Preliminary **Inventory File**

→ data preservation options and cost assessment

C - Decision on continuation of the data preservation in a dedicated Steering Committee (scientific needs and acceptation of the participation at the implementation phase, approval of financial and technical resources& schedule)

D - Implementation of the sustainability process data

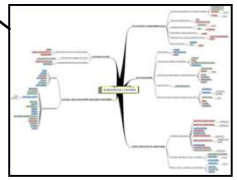
Recovery & data preparation, data archiving, metadata creation, data dictionary, data access criteria, data access (pblc,prvt), fmt conversion, data knowledge (PI)

final Inventory File

SERAD PILOT MISSIONS

In a preliminary work, about 180 missions/experiments (old,on going, future) with CNES participation have been inventoried in 4 thematics :

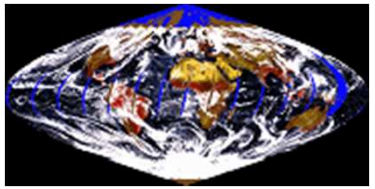
Sciences of the Universe / Earth Sciences / Life Sciences / Material Sciences

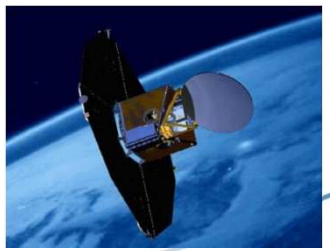


In order to improve and test data policy, the data management and to qualify tools, pilot missions have been chose

One by thematic :

- Polder for Earth Observation,
- Odin for Sciences of the Universe
- Mephisto for Microgravity.





FUNCTIONNAL DEFINITION

- Inventory function:

Provides an exhaustive inventory of the candidate data for referencing. The goal is to have an exhaustive list of CNES missions, and to be able to decide what to do with the datasets.

- Referencing function:

Allows the user to search data of interest through the discovery metadata, using criteria (interests) and keywords. Returns to the user the metadata for its selection.

- Archiving and ingestion functions

Allows archiving data from a producer and ingestion of metadata access in the SIPAD-NG catalogue.

- Access function

Allows users 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).

SERAD Fonctionalities

Missions

Orphan DATA
Project end of life

1

Referenc… …nventory projects
Keys word …
Refere…

Referencing activities

Survey activities

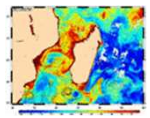
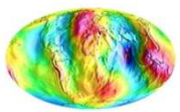
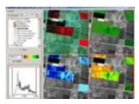
SERAD …IBOARD
CNES P…
Surveys … time)
Mana… …xploitation

Users

Long term A…
Data dictionar… …D

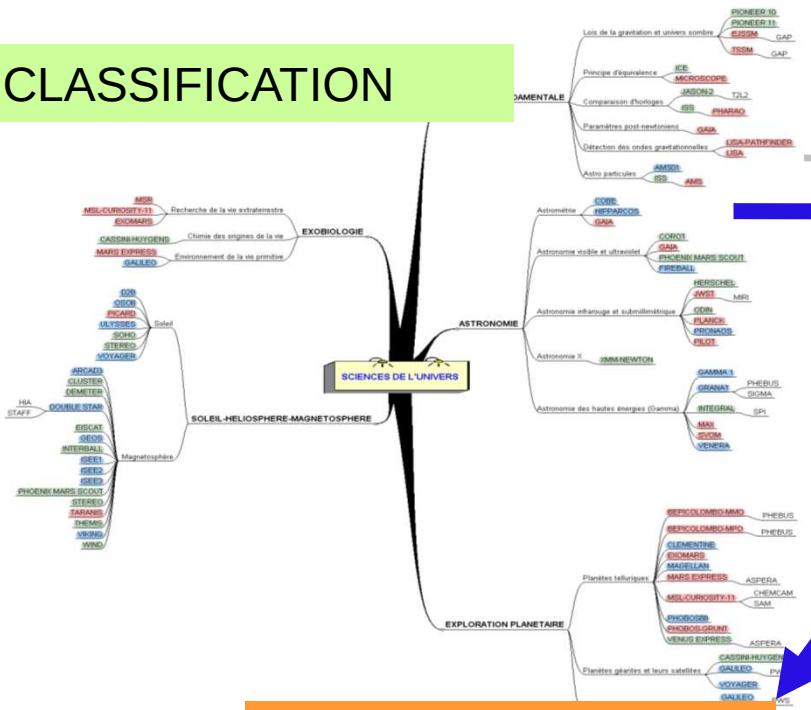
Long term Archive activities

Data Centers

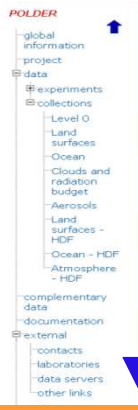

Long term archive
DSI STAF
pooled means

CLASSIFICATION



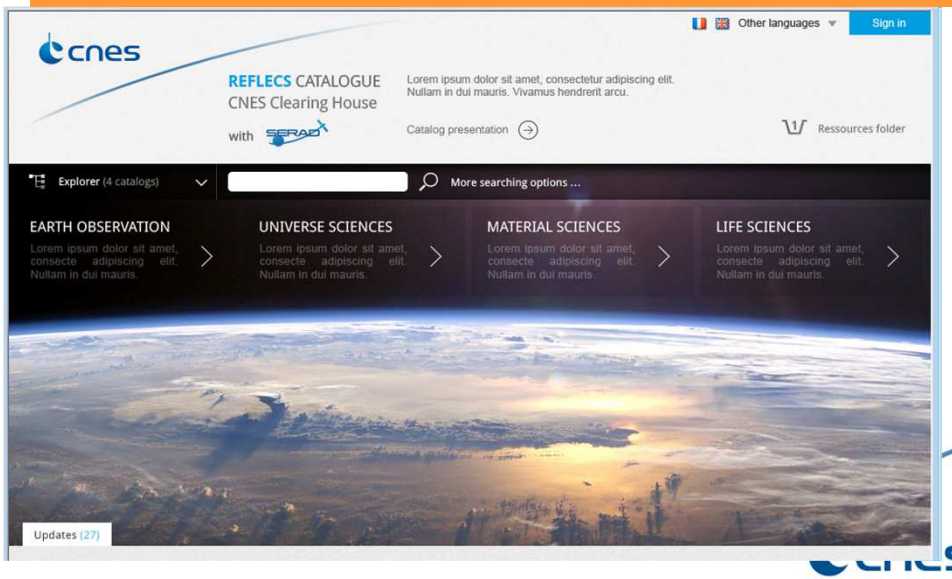
Inventory and referencing records
(metadata for discovery, key words)
ISO 19115 - INSPIRE

INVENTORY FILES



DASHBOARD

REFERENCING & ARCHIVE TOOLS



SERAD TOOLS

Generic - REFLECS CLEARING HOUSE : **(ISO, INSPIRE, CSW)** tools dedicated to dataset discovery service

Generic - BEST (OAIS, RNC): tools for formally describing the space data, validating, simulating data according to their formal description

Generic - SIPAD NG (OAIS): (System Information for Data Presevation and Access): implements the Data Management, Data ingestion and Data Access functions of OAIS. This configurable system is a standard product used for different data distribution centers

Shared - STAF : (Files Transfer and Archiving System) : implements the Storage function of **OAIS** for all the missions archived at CNES,CNES libraries tapes (SL8500) have just been replaced and are able to be increased up to 10 PB.

Specific- SERAD DASHBOARD (OAIS) : tool dedicated to manage SERAD activities (inventory, referency, archive)

The strategy for the architecture of SERAD archive tools is to be conform to the OAIS functional Model and to use CNES generic Products and Services

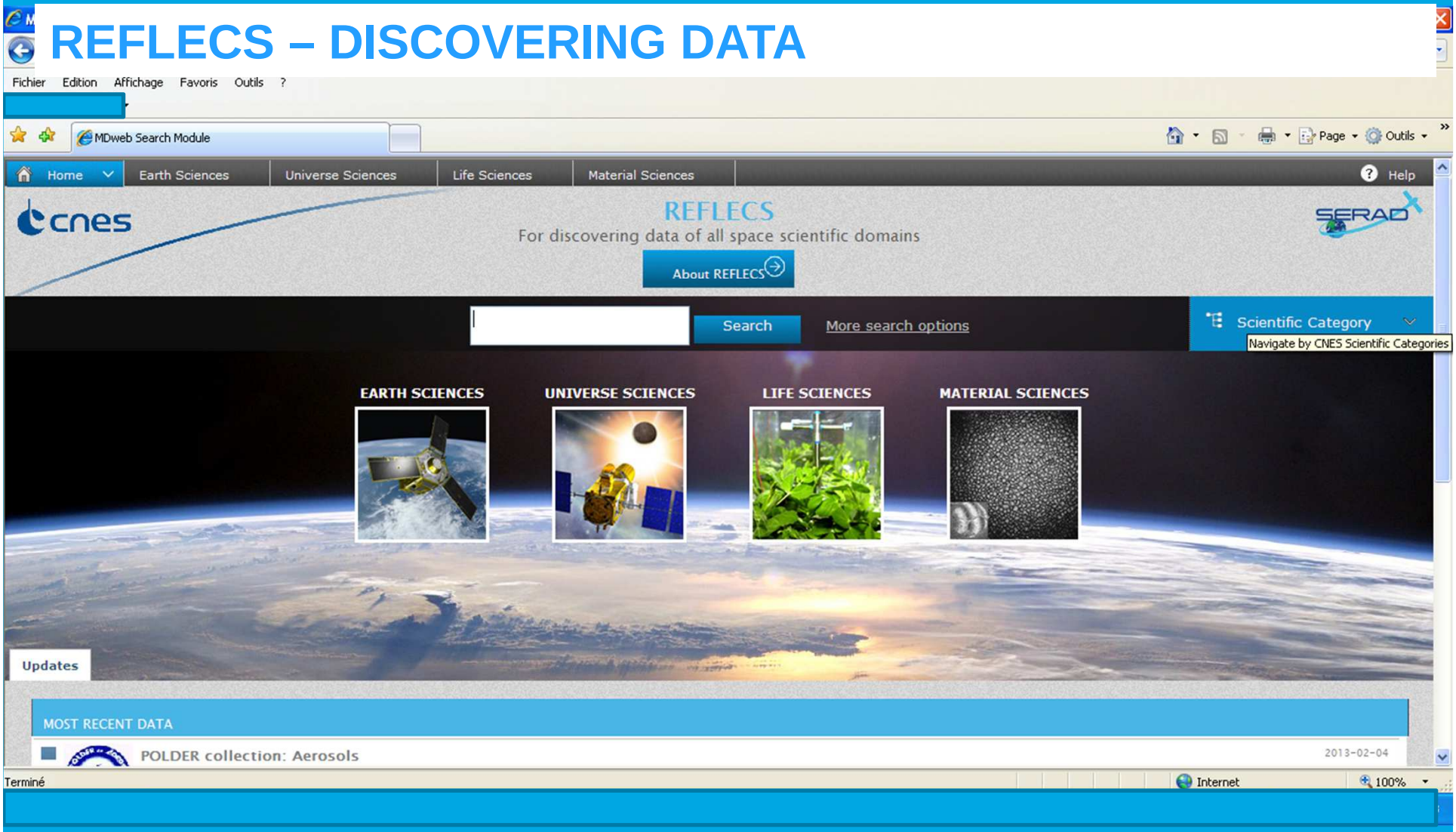
SERAD METADATA MODEL

For Clearing House REFLECS is based on :

- Classification
  - define the centers of interest which are representative, discriminating and consistent to the terminology of the future users.
  - precise classification job of all the science thematics. Help by CNES thematic responsible.

- Thesaurus
  - have an ordered list of keywords recognized by communities of users (common language)
  - in parallel with the classification of science thematics.

REFLECS – DISCOVERING DATA



Summary

Need to define and implement the **Data policy** and **« Best practices »** from start project

The growing volume agencies will **impose a cost reduction, an appropriate organizational structure**

Generics tools and shared resources are significant to reduce costs in the long term period

All **project members** are concerned with the preservation of data: scientific team, technical staff, layers, managers :

Thank you for your attention

martine.larroque@cnes.fr