wePreserve

The ESA SCIENTIFIC TESTBED in CASPAR

S. ALBANI & L. FUSCO *(ESA)*

Sergio.Albani@esa.int, Luigi.Fusco@esa.int

DPE, Planets and CASPAR
Third Annual Conference
Nice
29/30 October 2008

Information Society

digital preservation europe

planets

CASPAR

wePreserve

SUMMARY

- ESA EO data handling
- ESA data: benefits, risks and policies
- The ESA scientific testbed in CASPAR
- Conclusions

Information Society

digital preservation Europe

planets

CASPAR

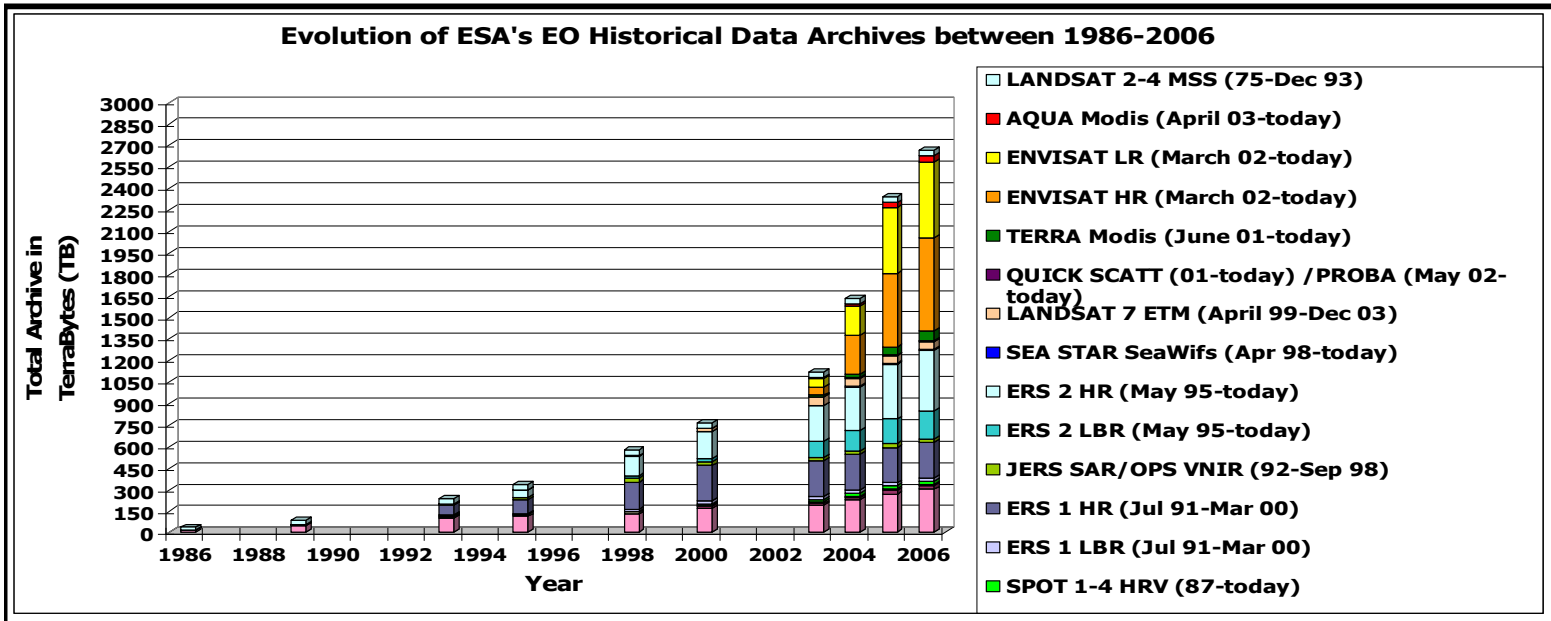
ESA EO Data Handling

- **ESA**, through worldwide receiving ground stations, acquires data from Earth Observation (EO) satellites and archives/processes them at Processing and Archiving Facilities.

- ESA-ESRIN is the **largest European EO data provider** and operates as the reference European centre for EO Payload Data Exploitation.

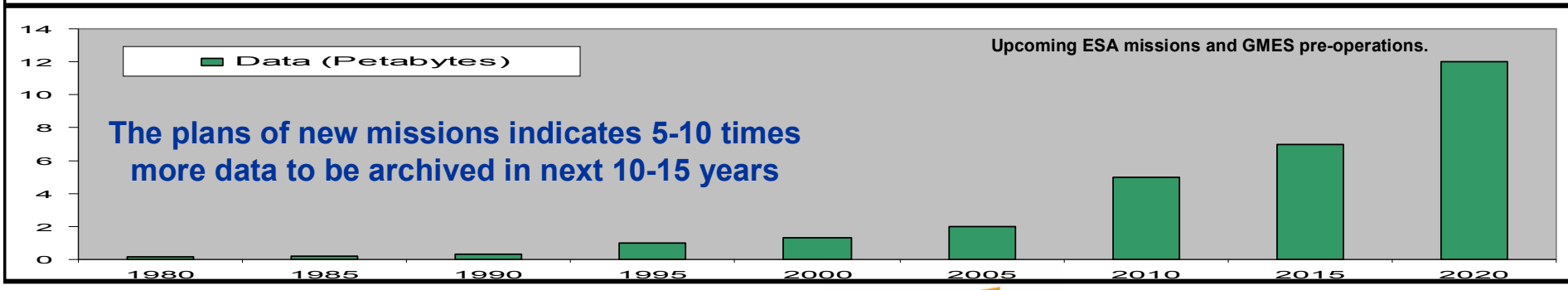
- At present, several thousands ESA worldwide users have online access to EO missions related metadata (10 million references), **data (about 3 PB)** and derived information.

The EO data avalanche...



ESA EO data archives include data from ESA missions (ERS and ENVISAT) and Third Party missions (Landsat, SPOT, ALOS, TIROS, AQUA&TERRA, etc.).

The evolution follows a similar trend for all national archives.



The plans of new missions indicates 5-10 times more data to be archived in next 10-15 years

Benefits, risks and policies (1)

- **EO archives ensure a global coverage of the Earth** with the following important characteristics:
  - multi-sensor data (from optical to active radar sensors);
  - long series (time-span that extends from a few years to decades);
  - variable geometrical resolution (from few meters to few hundreds meters);
  - variable geographical coverage (local, regional, global);
  - variable temporal resolution (from few days up to months).
- **The requirements for accessing ESA historical archives is strongly increased** in the last ten years and the trend is likely to increase in the future mainly for long term science and environmental monitoring.
- Therefore, the prospect of losing the digital records of science (and with the specific unique data, information and publications managed by ESA) is very alarming.

Benefits, risks and policies (2)

- **We have to preserve data** against changes in:
  - hardware,
  - software,
  - environment,
  - knowledge base of the scientific community…
- But current funds for ESA missions cover preservation and access to data for a baseline period limited to **only 10 years after acquisition!**
- The issues concern:
  - **type and amount of data to be preserved;**
  - **location of archives and their replication for security reasons;**
  - **technical choices (e.g. formats, media);**
  - **availability of adequate funds.**
- Decisions has to be taken in coordination with other data owners and with the support/advice of the user community.

Benefits, risks and policies (3)

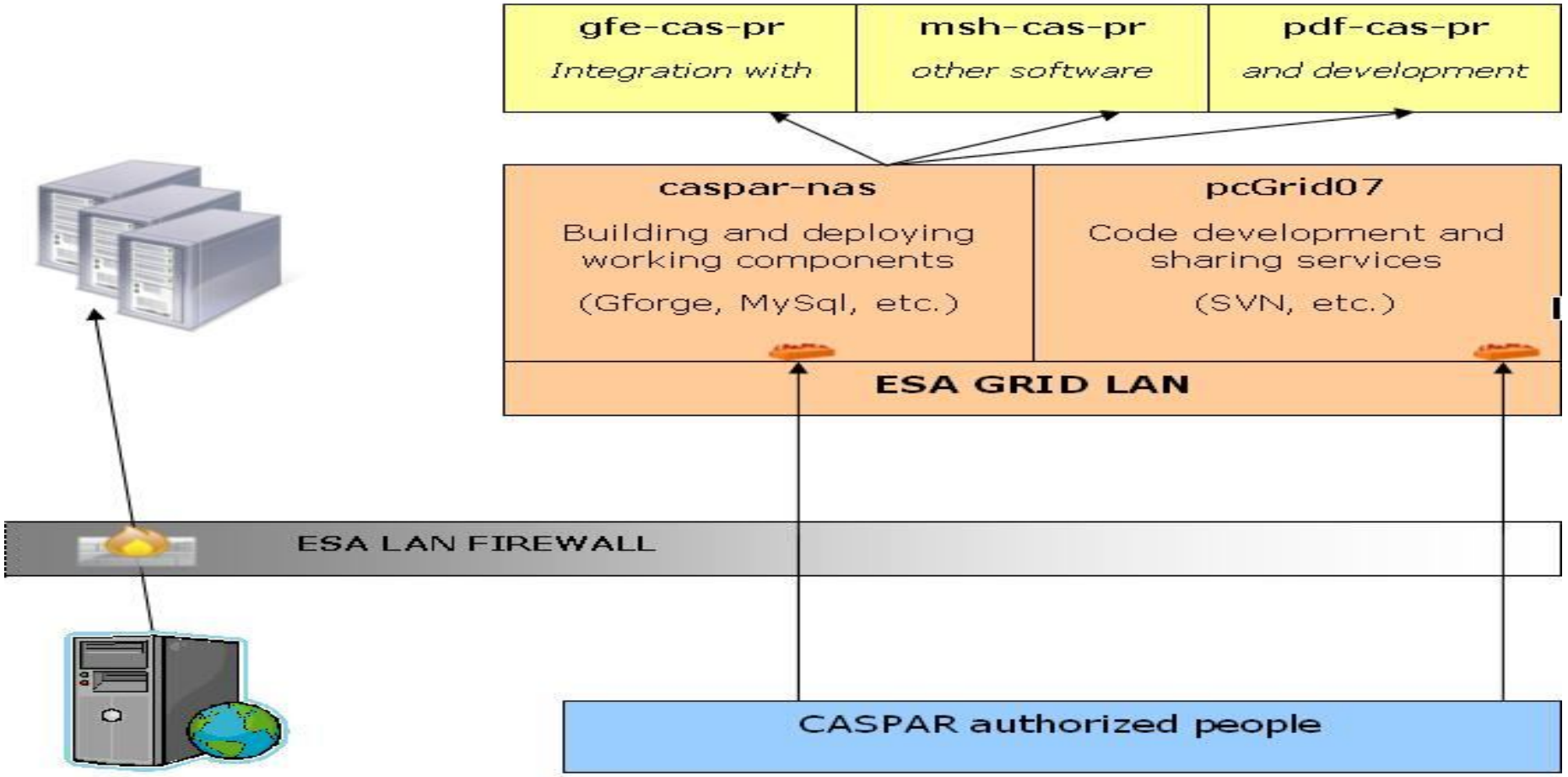
- So ESA is currently proposing to Member States to establish a viable **European-wide infrastructure for permanent access to the records of science** (data and publications) by the definition of:
  - an overall strategy for the EO data long term preservation;
  - an harmonized European archives management policy for ESA and Member States national EO data holdings .
- Meanwhile ESA:
  - is active in developing appropriate techniques and strategies by promoting and participating to projects related to long term data preservation;
  - is participating to the CASPAR project playing the role of both **user and infrastructure provider for the scientific data testbed** using new proposed standards, specialized infrastructures and open GRID environment.

ESA TESTBED ACTIVITIES

The ESA testbed is covering:

- the setup of the framework in ESA-ESRIN;
- the definition and collection of a significant sample of a whole processing chain dataset (viewers, converters, processors and data of different level, revision and format);
  - the conversion of the data from the native format to a OAIS compliant format;
  - the generation of appropriate Representation Information, Descriptive Information, Knowledge Modules and Scientific Community profiles;
  - the analysis of ontologies to describe and preserve scientific workflows (e.g. the applicability of CIDOC CRM on scientific data);
  - the ingestion of data in the CASPAR system;
  - the coping with some long term data preservation problems by using the CASPAR components, methodology and tools.

Testbed: HW deployment in ESRIN

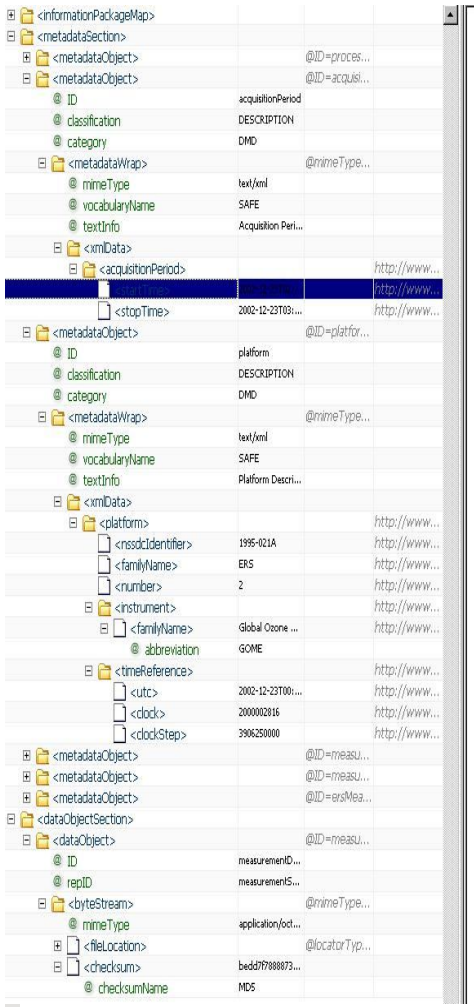


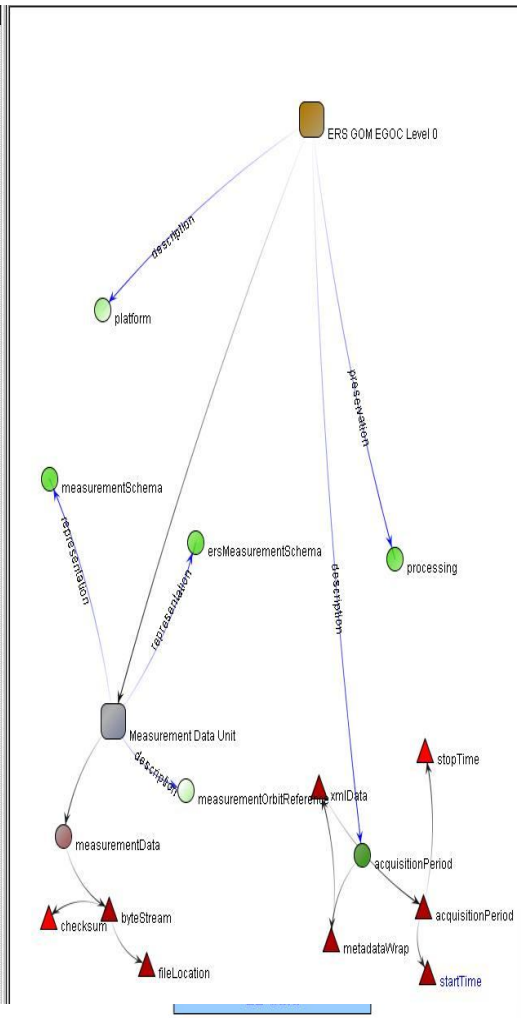
Testbed: the GOME dataset (1)

- The ESA selected dataset as a suitable demonstration case in the framework of the scientific testbed of the CASPAR project consists of data from **GOME (Global Ozone Monitoring Experiment)**, a sensor on board the ESA ERS-2 (European Remote Sensing) satellite.

- The GOME dataset:
  - has a big total amount of information distributed with a high level of complexity;
  - is unique because it provides more than 11 years global worldwide coverage;
  - is very important for the scientific community and the Principal Investigators (e.g. KNMI and DLR) that on a routine basis receive GOME data for their research projects (e.g. concerning ozone depletion or climate change);
  - is just a test-case because similar issues involve many other Earth Observation instrument datasets.

Testbed: the GOME dataset (2)





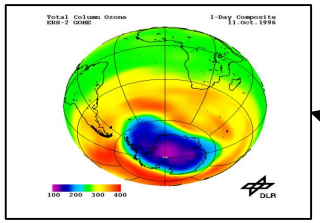
- Level 0: raw data
- Level processors
- Auxiliary data
- Documents and methods
- Level 1: radiances/reflectances
- Level 2: geophysical data as trace gas amounts
- Level 3: a mosaic composed by several level 2 data with interpolation of data values to fill the satellite gaps
- Data viewers
- Examples of GOME science applications
- Format converters

ESA TESTBED ACTIVITIES

The ESA testbed is covering:

- the setup of the framework in ESA-ESRIN;
- the definition and collection of a significant sample of a whole processing chain dataset (viewers, converters, processors and data of different level, revision and format);
- the conversion of the data from the native format to a OAIS compliant format;
- the generation of appropriate Representation Information, Description Information, Knowledge Modules and Scientific Community profiles;
- the analysis of ontologies to describe and preserve scientific workflows (e.g. the applicability of CIDOC CRM on scientific data);
- the ingestion of data in the CASPAR system;
- the coping with some long term data preservation problems by using the CASPAR components, methodology and tools.

Testbed: going to OAIS



The filename itself.

Empty.

A Cyclical Redundancy Check (CRC) code for a file.

The principal investigator who recorded the data and the information concerning its storage, handling and migration.

Content Information

Data Object

Representation Information

Preservation Description Information

Reference

Provenance

Context

Fixity

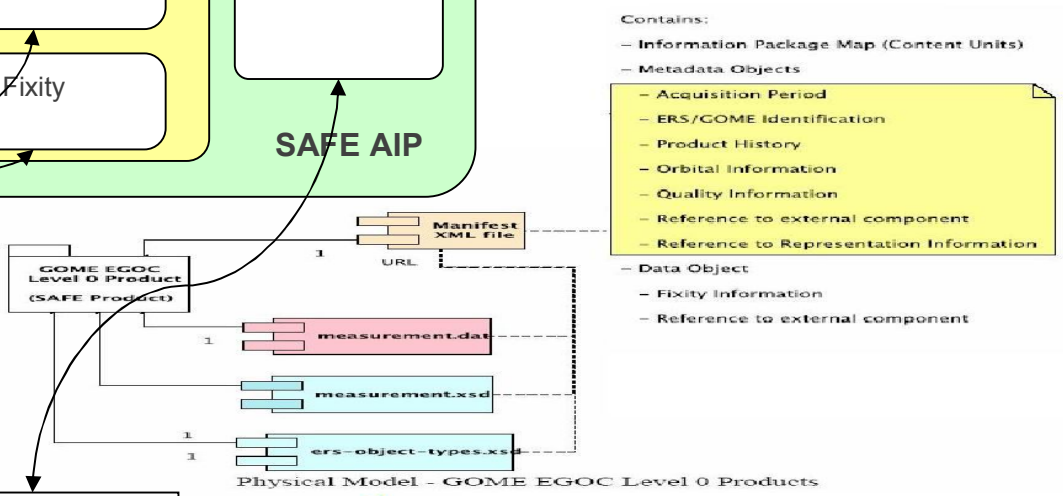
Descriptive Information

Packaging Information

SAFE AIP

The RepInfo provided are contained in the manifest file and in the schemas.

Metadata is extracted by the product and contained in the SAFE manifest.



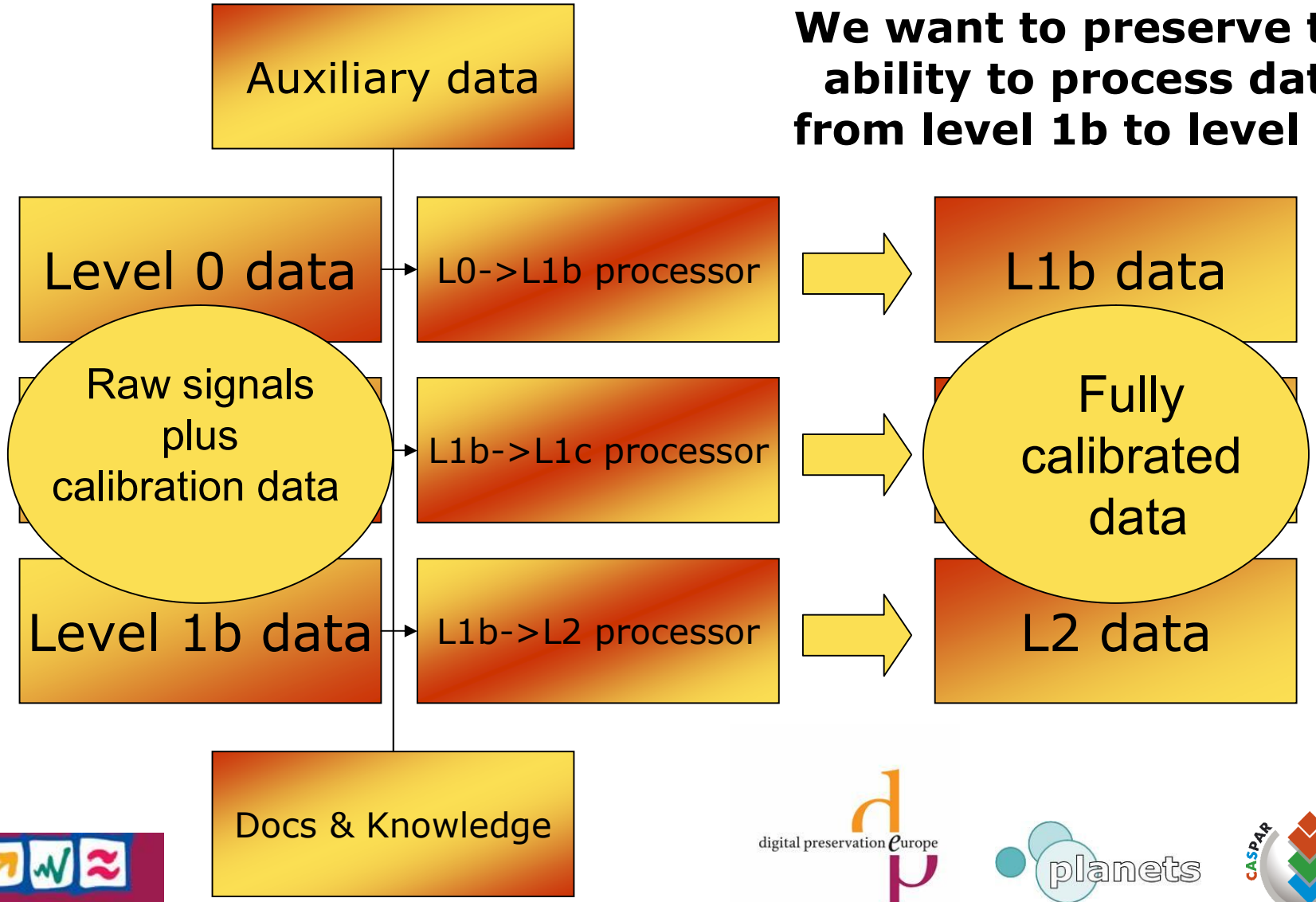
No packaging restrictions.

ESA TESTBED ACTIVITIES

The ESA testbed is covering:

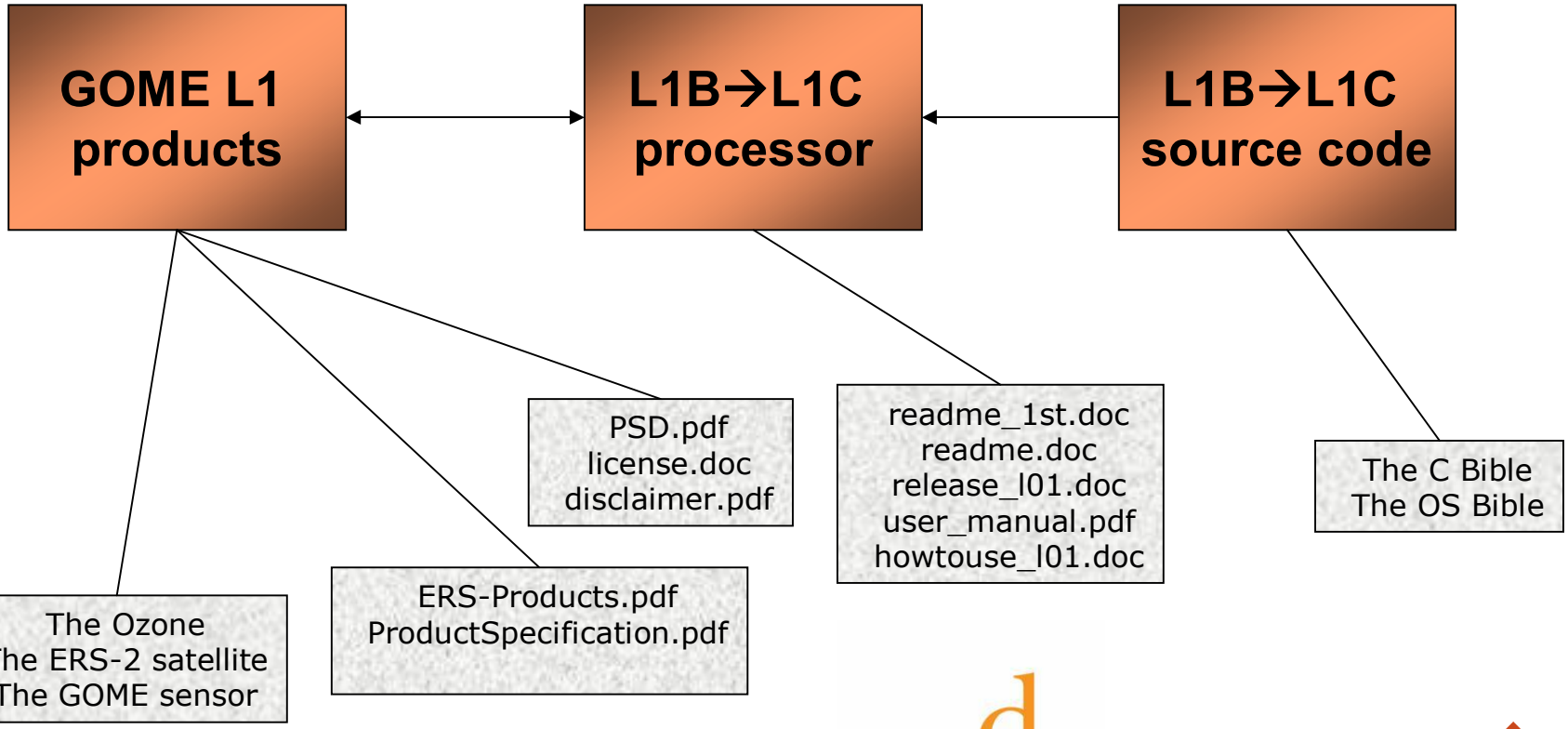
- the setup of the framework in ESA-ESRIN;
- the definition and collection of a significant sample of a whole processing chain dataset (viewers, converters, processors and data of different level, revision and format);
- the conversion of the data from the native format to a OAIS compliant format;
- the generation of appropriate Representation Information, Descriptive Information, Knowledge Modules and Scientific Community profiles;
- the analysis of ontologies to describe and preserve scientific workflows (e.g. the applicability of CIDOC CRM on scientific data);
- the ingestion of data in the CASPAR system;
- the coping with some long term data preservation problems by using the CASPAR components, methodology and tools.

Testbed: preservation scenario (1)

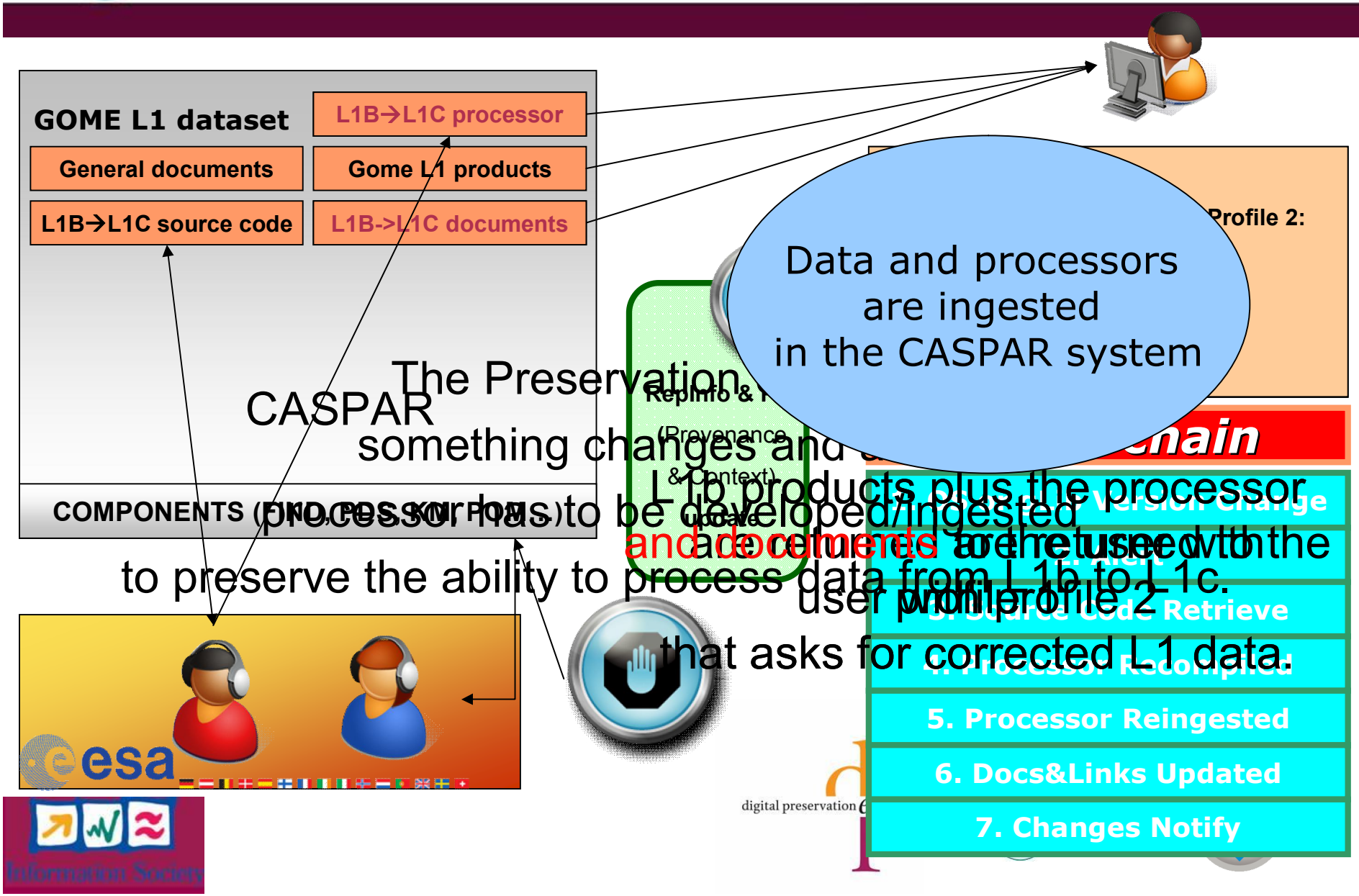


Testbed: preservation scenario (2)

We have to preserve the GOME L1 products, processors, reference manuals, etc.



Testbed: preservation scenario (3)



CONCLUSIONS

- The preservation of EO data is of vital importance for the scientific and operational user communities.
- ESA is pursuing the objective to ensure the perpetual preservation of these data in coordination with institutions of its member states.
- ESA data preservation initiatives will benefit of the results of CASPAR (and other similar EU sponsored projects) by adopting, when applicable, technical solutions and procedures developed in the framework of these cooperative projects.

wePreserve

THANKS!