



An actual use case: the ESA SCIENTIFIC TESTBED in CASPAR

S. ALBANI (ESA)
Sergio.Albani@esa.int

CASPAR Training Day for the Scientific Domain, Rome, 12 January 2009



SUMMARY

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



ESA EO Data Handling

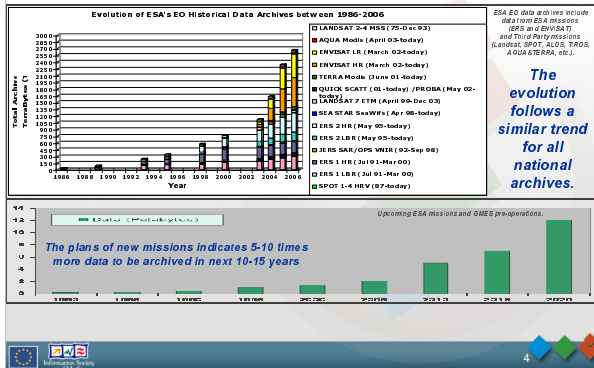
• ESA, through worldwide receiving ground stations, acquires data from Earth Observation (EO) satellites and archives/processes them at Processing and Archiving Facilities.

• 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...



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 tested using new proposed standards, specialized infrastructures and open GRID environment.

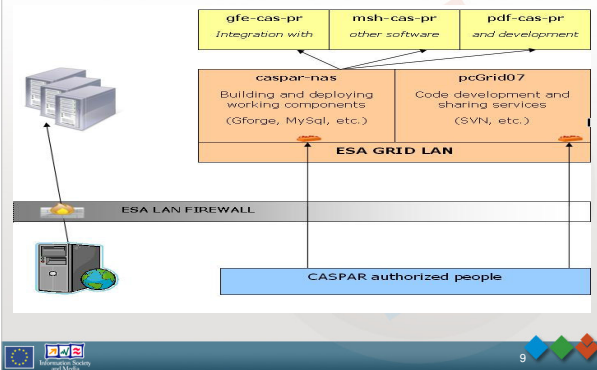
ESA TESTBED ACTIVITIES

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.

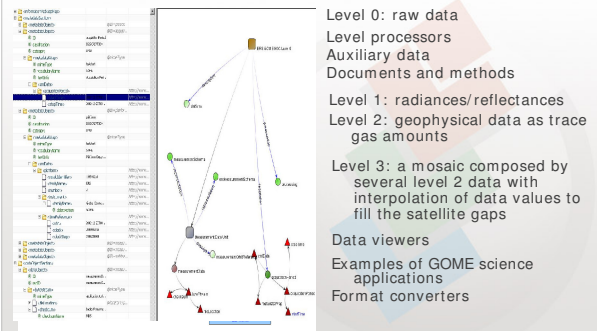
Testbed: HW deployment in ESRIN



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)



ESA TESTBED ACTIVITIES

The ESA testbed is covering:

- 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.

The ESA's SAFE format and MMFI

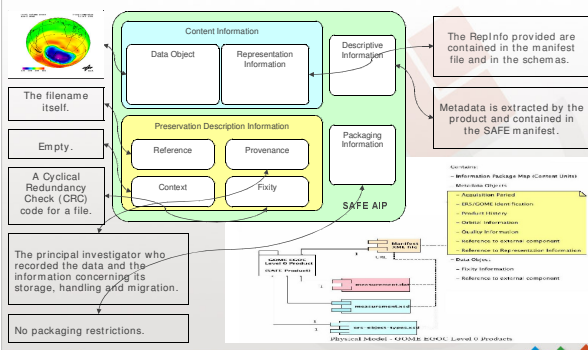
The **Multi Mission Facility Infrastructure** distributed system architecture:

- responds to ESA strategy for evolution of Earth Observation missions;
- has the goal to define a harmonized European ground segment infrastructure;
- **is based on the ISO 14721:2003 OAIS standard;**
- **provides producer oriented services for data archiving, data retrieval and processing management;**
- provides consumer oriented services in support of the ESA central infrastructure services;
- is utilized in several of ESA present operational facilities;
- uses a specialised form of the CCSDS XFDU (XML Formatted Data Units) standard for content packaging (SAFE).

The Standard Archive Format for Europe:

- has been developed by ESA in the framework of its EO ground segment activities;
- **acts as standard format for archiving and conveying data within ESA EO archiving facilities and potentially with the cooperating agencies;**
- **conforms to the ISO 14721:2003 OAIS (Open Archival Information System) reference model;**
- **is based on the XFDU standard;**
- offers a single framework for packaging a large variety of information;
- has the goal of preserving the archived data for a long-term:
 - facilitating the conversion into different formats,
 - simplifying the extraction from the archive,
 - enhancing their utilization by end-users and/or processing systems.

GOING TO THE OAIS

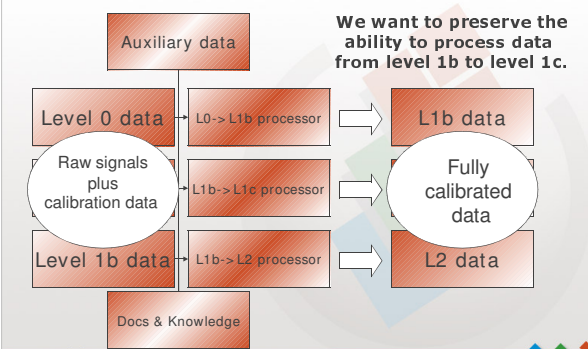


ESA TESTBED ACTIVITIES

The ESA testbed is covering:

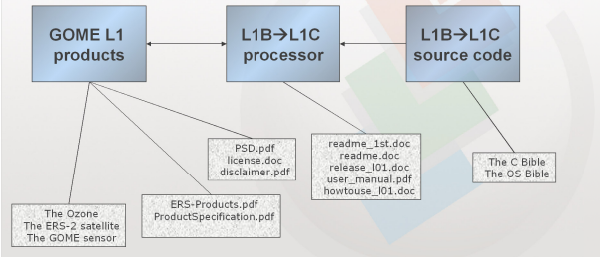
- the setup of the framework in ESA-ESRN;
- the definition and selection of a significant sample of a whole processing chain (viewers, converters, processors and data of different level, revision and format);
- the conversion of the data from the native format to a OAIS compliant format;
- the generation of appropriate Representation Information, Descriptive Information, Knowledge Module and Scientific Community profiles;
- the mapping of knowledge to describe and preserve scientific activities (e.g. the applicability of CDDC CRM for 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)

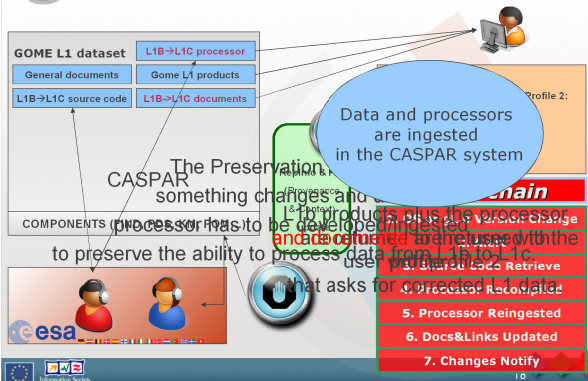


Testbed: preservation scenario (2)

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.



THANKS!