Long Term Preservation of Earth Observation Data

QA4EO Workshop
RAL, October 18-20th 2011

Mirko Albani and Bojan Bojkov* (ESA/ESRIN)
Outline

- Earth Observation data preservation: the need and the European challenge
- ESA LTDP Preliminary Programme 2009-2011: Status and Achievements
- LTDP and QA4EO
- Conclusions
Outline

• Earth Observation data preservation: the need and the European challenge

• ESA LTDP Preliminary Programme 2009-2011: Status and Achievements

• LTDP and QA4EO

• Conclusions
We have now reached about 20 years of continuous measurements from space for many geophysical parameters (even 30 years for some parameters)

→ Earth Observation data are becoming an essential tool for Earth Science

→ Long term series of data needed for several applications in different Earth Science domains
The preservation of EO data (the “bytes”) is useless without the preservation of the knowledge associated with the data.
Data Preservation Key Aspects:

• Preservation of all EO data in the different categories (e.g. SAR, optical, etc.) but also of all associated information (documentation, CAL/VAL databases, algorithms, etc.).

• Completeness and coherency among all elements to be preserved to ensure present and future exploitability.

• Data Quality, Context and Provenance (documented).

• Maintenance of capabilities to (re-)generate data products.

• Harmonized data and information accessibility.

• Coordinated and coherent approach in Europe.
The challenge: EO Archives Data volume and long term sustainability

ESA Data Archives (through 2010) and projections through 2020

Projections

Difficult to locate long term funding and no programmatic approach in place for LTDP: ESA data preservation mandate is for 10 years after end of mission (e.g. no mandate to preserve ERS-1 data anymore).
There is still a lot to discover in the EO data archives

A recent example of unexpected result with the 20 years data archive:

“Monitoring the South Atlantic Anomaly using ATSR instrument series”, S. Casadio & O. Arino

[paper published in COSPAR Advance in Space Research]
Outline

- Earth Observation data preservation: the need and the European challenge
- ESA LTDP Preliminary Programme 2009-2011: Status and Achievements
- LTDP and QA4EO
- Conclusions
LTDP was approved in 2008 (3-year funding) in order to:

1. Prevent ESA and ESA Third Party Mission (TPM) EO data loss → Implementation activities

2. Start the implementation of the LTDP infrastructure needed for data and knowledge management → Evolution activities

3. Identify the minimum “dataset content” to be preserved and define the LTDP architecture for efficient preservation and accessibility of the ESA EO data and associated knowledge in the long term → Analysis and studies

4. Define & reinforce a common approach for LTDP in Europe → European LTDP framework coordination

5. Prepare the LTDP Programme proposal for the period beyond 2012. → LTDP programme preparation
Implementation in ESA archives: Integrity of archived data

1. Audit the ESA archive consistency covering data from the historical Third Party Mission since late 70’s: on going, inventory consolidated.
   - **Ingestion of old media** into automated tape libraries for higher priority datasets (CCI initiative): on going for ERS-1/2, LANDSAT, JERS-1.
   - **Data recovery** from old corrupted tapes and ingestion: completed for ERS, Landsat TM/ETM+/MSS, Terra-Aqua MODIS, SPOT HRV, JERS-1.
   - **Recovery/consolidation** of all Auxiliary files for ERS-1/2.

⇒ **Status of all ESA data identified (and ingested) no longer at risk of loss.**

2. Archive technology upgrade needed at all ESA facilities to improve (cost-effective) data preservation and access.

3. Analysis for archives rationalization (with the goal of two complete and identical copies to be stored in two different locations).
Implementation activities: Knowledge preservation beyond data (e.g. documents)

1. Audit the available data and documentation w.r.t. “Dataset Content” for preservation:
   ⇒ “Dataset Content” for preservation defined and consolidated through instantiation with some missions.
   ⇒ Knowledge for these missions recovered (at least partially) in addition to the data and available for Climate Change Initiative and other uses.
   ⇒ Recovery begun for ERS-1/2, NOAA AVHRR and LANDSAT series.

2. Completing the audit and recovery for all ESA and ESA TPM managed historical missions will be undertaken in the follow-on LTDP Programme activities.
Evolution activities: Data access and processors

1. Archived data on-line
   - Progressive provision of on-line access to archived data (e.g. SPOT, ERS, ENVISAT).
   - Enhancement and extension to additional missions (e.g. LANDSAT, Seawifs, JERS) and security consolidation.

⇒ Ensure integrity and availability on-line of data/products from all ESA missions and several TPMs allowing easy access to users and in support of the CCI initiative.

2. Processing capacity preservation
   - Virtualization of a subset of processors and implementation of prototype virtualized environment.
   - Processors independent on the underlying SW/HW reducing maintenance costs.

⇒ Facilitating the preservation of the processing and reprocessing capacity.

On-line accessibility of data and maintenance of the processing/reprocessing capacity is fundamental to support EO data users and exploitation programmes.
Outline

• Earth Observation data preservation: the need and the European challenge

• ESA LTDP Preliminary Programme 2009-2011: Status and Achievements

• LTDP and QA4EO

• Conclusions
ESA is coordinating the LTDP activities in the Earth Observation domain with European partners through the LTDP WG formed within the Ground Segment Coordination Body (GSCB):

1. Started the creation of the European LTDP Framework in Earth Observation, implementing the basic rules, “LTDP Common Guidelines”.

2. Defined the initial data set to be preserved, including the related glossary.

3. Guaranteed the information flow through workshops, web sites, participation to conferences and LTDP related events.
LTDP Common Guidelines

• Basic “guidelines” of the European LTDP Framework in Earth Observation:
  ➢ Reflecting the “consensus” and “best practice” of the European EO Data providers.
  ➢ Reviewed at GEO, CEOS and with NASA bilaterally.
  ➢ [http://earth.esa.int/gscb/ltdp/EuropeanLTDPCo..._Issue1.1.pdf](http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_Issue1.1.pdf)

• Addresses preservation, processing, access and exploitation of data.

• Application by EO space data owners and archive holders fundamental to preserve EO space data for the benefit of European users.
  – Could follow a step-wise approach starting with a partial adherence (three different levels of adherence have been defined) and full adherence to be pursued in the long term.

• GEO EO LTDP Guidelines issued in November 2010 (Issue 1.0, aligned with European ones) in the framework of GEO Sub-task DA-09-01C.
LTDP Common Guidelines Structure

- Structured as a set of eight main themes each with associated a “Definition”, a “Guiding Principle” and a set of “Key guidelines”:
  1. Preserved data set composition
  2. Archive operations and organization
  3. Archive security
  4. Data ingestion
  5. Archive maintenance
  6. Data access and interoperability
  7. Data exploitation and re-processing
  8. Data appraisal and purge prevention

- Each key guideline can have associated more detailed technical procedures / methodologies or standards necessary for practical implementation.

- Guidelines and procedures might evolve with time to incorporate the results of LTDP related activities (e.g. outcome of cooperation among LTDP actors).
QA4EO and LTDP: THEME 5 – Archives maintenance

• Definition (Perimeter of the theme)

• Guiding Principle (Major objective and approach)
  Earth Observation space data owners and archive holders should design a maintenance scheme for their archives to guarantee the integrity of the archived data.

• Key Guidelines
  GUIDELINE 5.1 – Archived data refreshment - (Level A): Perform periodically a migration of the archived data (“media refreshment”) to the most adequate proven technology for data storage ensuring data access preservation[1].

  e.g. Technical procedure for data migration (e.g. from GEO task, CCSDS, CEOS WG)

[1] Currently data and system migrations are performed at least every five/six years.
Pre-flight & Commissioning examples:

- Information on the instrument (e.g. including the design) and its characterization documentation
- Pre-flight calibration procedures and data
- Processors (L0 to L2) documentation
- End-to-end uncertainty budget including applied models and auxiliary files
- Commissioning phase report
 QA4EO & LTDP (iii): Operational phase example

- Instrument performance (e.g. degradation of subsystems) throughout its lifetime
- Processors (L0 to L2) evolution with all relevant documentation with reprocessing history
- Calibration and validation procedures (both on-board and ground based) and relevant data
- Relevant spacecraft platform data
- Operational data quality control reports (product quality reports and mission performance reports)
- Anomaly reports (for spacecraft and ground segment)
Specific QA4EO activity within ESA

A dedicated ESA **General Study Programme ITT** for the:

**EVALUATION OF REQUIREMENTS ON DATA QUALITY INFORMATION IN RELATION THE LTDP GUIDELINES**

Not only EO missions but also ESA Science Missions: KO planned in Q1 2012
Outline

• Earth Observation data preservation: the need and the European challenge

• ESA LTDP Preliminary Programme 2009-2011: Status and Achievements

• LTDP and QA4EO

• Conclusions
Conclusions

• Needs and challenges related to LTDP of EO data are well identified

• EO LTDP Preliminary Programme is allowing the:
  – Implementation of essential activities at ESA facilities focusing on data preservation (e.g. ingestion in tape libraries), recovery (e.g. ERS mission) and access enhancement (i.e. on-line) prioritizing data required by CCI.
  – Definition of LTDP GS architecture for knowledge and data preservation and data recovery/rationalization plans for implementation in future.
  – Coordination of LTDP activities, with the involvement of all European data owners and archive holders, and LTDP guidelines consolidation and promotion.
Thank you!

Any question?

Contacts: Mirko.Albani@esa.int