

# **LONG TERM DATA PRESERVATION USER NEEDS WORKSHOP (LTDP/FIRST)**

ESA/ESRIN 21<sup>ST</sup> OCTOBER 2010

---

**WORKSHOP REPORT**

## CHANGE LOG


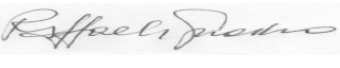
Reason for change	Issue	Date
First emission	1.0	09 November 2010

## CHANGE RECORD

Issue: 1.0

Reason for change	page(s)	Paragraph(s)
First emission	All	All

## APPROVALS

Role	Name	Signature
Writer	Barbara Polsinelli (Capgemini)	
Review and Approval	Raffaele Guarino (Capgemini)	
Authorisation	Mirko Albani (ESA)	

## Table of Contents

<b>1. ABSTRACT</b> .....	<b>4</b>
<b>2. ANNEXES</b> .....	<b>5</b>
<b>3. WORKSHOP</b> .....	<b>6</b>
3.1 OBJECTIVES.....	6
3.2 ORGANIZATION.....	7
3.3 PARTICIPATION.....	8
<b>4. OUTCOME</b> .....	<b>9</b>
4.1 LONG TERM DATA PRESERVATION STATUS.....	9
4.1.1 <i>FIRST (Definition of LTDP User Requirements and Preserved Data Set Composition) study</i> ..	9
4.1.2 <i>Comments and recommendations</i> .....	11
4.1.3 <i>Oceanography case</i> .....	13
4.1.4 <i>Cryosphere case</i> .....	14
4.1.5 <i>Geophysics case</i> .....	14
4.2 CONCLUSIONS AND NEXT STEPS.....	14
<b>5. ANNEX 1 – WORKSHOP AGENDA</b> .....	<b>16</b>
<b>6. ANNEX 2 – LIST OF PARTICIPANTS</b> .....	<b>18</b>

## 1. ABSTRACT

On 21<sup>st</sup> October 2010 the **Long Term Data Preservation User Needs workshop (LTDP/FIRST)** was held at ESA/ESRIN premises.

The workshop was addressed to Earth Science and Earth Observation scientific data users, owners and providers. It was devoted to present and compare the results achieved so far by LTDP/FIRST Project in analysing user requirements coming from data users.

Preservation in the long term of Earth Science data is a major challenge in Europe: LTDP cooperation activities aimed at guaranteeing the preservation of European data are ongoing in the Earth Observation domain under ESA coordination. FIRST is considered as an important step for the extension of the current European cooperation framework from the Earth Observation domain to the wider Earth Science concept. The goals of LTDP/FIRST Project are:

- Collection and definition of requirements and needs from Earth Science users in terms of Long Term Preservation of Earth Science data, products and related information.
- Definition of the composition of the Earth Science data set that should be preserved in the long term.
- Preliminary assessment of the current capabilities of European archives and provision of recommendations for the set-up of a European LTDP Framework.

The workshop was aimed at confirming and improving the understanding of Earth Science communities needs in term of data necessary to support research and applications. The workshop was also a substantial input for the consolidation of the European Earth Observation LTDP framework, aiming at guaranteeing the preservation of Earth Observation data for the benefits of all users and in general of European and Canadian citizens, and for its extension to the wider Earth Science domain.

## 2. ANNEXES

Annexes are:

- Annex 1 : Workshop agenda
- Annex 2 : List of participants

Presentations held at the workshop can be found on the GSCB/LTDP web-site:

<http://earth.esa.int/gscb/ldp/agenda-workshop2010.html>

## 3. LTDP/FIRST USER WORKSHOP

### 3.1 Objectives

The long term preservation of Earth Observation data is a major issue today as monitoring of global change processes has led to increasing demand for long-term time series of data spanning 20 years or more also in support to international initiatives such for example the United Nations Framework Convention on Climate Change (UNFCCC), the ESA Climate Change Initiative (CCI) and the GMES program. The large amount of new Earth Observation missions upcoming in the next years will lead to a major increase of EO space data volumes and this fact, together with the increased demands from the user community, marks a challenge for Earth Observation satellite operators, Space Agencies and EO space data providers regarding coherent data preservation and optimum availability and accessibility of the different data products. Traditionally in Europe, there has been poor cooperation in this field with no common approach for long term preservation and access to EO space data and single organizations have difficulties to afford data preservation in the long term that calls for the need of optimising costs and efforts, identifying commonalities.

ESA started at the beginning of 2008 a coordination activity with the goal to set-up a cooperative and harmonized collective approach on Long Term Data Preservation (LTDP) in Europe (i.e. a European EO LTDP Framework) to optimize European efforts in the LTDP field and to ultimately result in the preservation of the complete European EO space data set for the benefit of all European users and with a reduction of overall costs. ESA and its partners, through the Ground Segment Coordination Body (GSCB) LTDP working group, have already published, as a first concrete step for the set-up of this framework, an LTDP guidelines document publicly reviewed by all European data owners and available at the link: [http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines\\_Issue1.0.pdf](http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_Issue1.0.pdf).

An LTDP Preparatory Programme was also approved by ESA member states in 2008 for the period 2009-2011 having the goal to implement the most urgent activities to guarantee the preservation in the long term of ESA managed Earth Observation data and to continue the ongoing LTDP cooperation activities aimed at setting up the European LTDP framework.

In this context ESA has started a project with an European company (Capgemini Italia S.p.A) named FIRST (deFinition of LTDP user Requirements and preservation data Set composition) aimed at providing a complete and detailed view and analysis of the Earth Science users requirements and needs in all fields of applications in terms of Long Term Preservation of Earth Science data, products and related information and at defining the composition of the Earth Science data set that should be preserved in the long term.

The aim of the **Long Term Data Preservation User Needs workshop (LTDP/FIRST)** organized at ESRIN by the ESA Earth Observation Ground Segment department (EOP-G) with the support of Capgemini Italia SpA was to:

- Share and provide an overview of Earth Science users needs and requirements for the data preservation captured in the frame of the FIRST project.
- Integrate additional user needs and feedbacks to consolidate the Earth Science users' community requirements.
- Present the new ESA missions 'data policy to the Earth Science users' community.
- Present the LTDP activities and the future vision.

The workshop provided a forum for Earth Science users to present their needs and expectations in terms of long term data preservation and access and to provide feedback on the ongoing LTDP activities.

## 3.2 Organization

Workshop invitation was sent two months in advance with respect to the workshop planned date. The workshop was organized to run on one day (the detailed agenda is in Annex 1) and was split into the following sessions:

- 1) **Introduction:** ESA Welcome and introduction, objectives, work plan.
- 2) **Session 1: Long Term Data Preservation and FIRST activities and results.** The session was aimed at presenting the status and results of the LTDP and FIRST activities and included presentations on:
  - a. European LTDP activities and guidelines
  - b. LTDP activities and needs in GEO
  - c. QA4EO, preservation requirements for quality assurance
  - d. EC and European Scientific Data e-Infrastructures: an overview of the latest developments and future perspectives
  - e. FIRST Project preliminary results
- 3) **Session 2: Long Term Data Preservation User Needs.** The session was aimed at presenting the needs from each Earth Science fields taking in consideration some applicative cases and experience of intergovernmental organizations and research centers. The session included presentations on:
  - a. An EC research project: ERA-ORA
  - b. Needs from science: an **Oceanography** case
  - c. Needs from science: a **Cal/Val – Radiometry** case
  - d. **Climate Change Initiative** case
  - e. Environmental Science Information Requirements
  - f. **GMES** Services case: SAFER
  - g. Needs from science: the **DLR** case and experience
  - h. **ECMWF** case and experience
  - i. The New OGC Coverage Model and oits Services: Case and Perspectives
  - j. Needs from science: a **Cryosphere** case
  - k. Needs from science: a **Geophysics** case
  - l. **DFG** (German Research Foundation) case and experience
  - m. Evolution of **ESA** services
- 4) **Session 3: Wrap-up and Discussion.** The session was aimed at providing a short summary of the workshop results and promoting discussion to collect additional feed-back on current LTDP/FIRST activities and future vision.

During **Session 1** three flipcharts were introduced with the goal to capture feedbacks from the participants. On two flipcharts, were presented open points and we asked for ideas and contributions or comments about; one was open to capture opinions and ask questions, make suggestions about topics covered in the day freely. Post-it notes were delivered to participants that they could fill and they could attach as they liked.

Considering the interest about the theme and the number of received requests to participate the event it was possible to follow the workshop via living video stream at the following URL: [http://www.esa.int/esaEO/SEMADRVO1FG\\_index\\_0.html](http://www.esa.int/esaEO/SEMADRVO1FG_index_0.html)

### 3.3 Participation

The workshop was addressed to Earth Science and Earth Observation research communities and data users. A total of 27 organizations have been participating.

These organizations were:

- **International Organizations, Governmental Entities and Space Agencies:** CNES (French Space Agency), CSA (Canadian Space Agency), DLR (German Space Agency), European Commission (EC), European Centre for Medium-Range Weather Forecasts (ECMWF), European Space Agency (ESA), Finnish Meteorological Institute (FMI), GEO Secretariat.
- **Research Institute:** Applied Research in Geomatics, Atmosphere, Nature & Space (Argans Limited, UK), European Academy of Bozen/Bolzano (EURAC, IT), German Research Foundation (DFG, DE), National Institute of Geophysics and Volcanology (INGV, IT), Jacobs University (DE), Norwegian Meteorological Institute (Met.no, NO), Netherlands Institute for Space Research (SRON, NL), Tor Vergata University (IT), German Research Centre for Geosciences (GFZ, DE), Research Centre Institute for Coastal Research (GKSS, DE).
- **EO Data owners and providers:** Euromap GmbH, Infoterra Ltd UK, STFC, NERC Earth Observation Data Centre UK, German Remote Sensing Data Center (DLR-DFD).
- **Other Ground segment actors:** Advanced Computer Systems (ACS) SpA, Capgemini, GMV Aerospace and Defence SA (GMV, E), ROVSING A/S (DK), Swedish Space Corporation (SSC, S).

The detailed reference list of participants is provided as Annex 2 to this document.

Some registered people were not able to attend the workshop due to problems in air traffic in France with some side effects in other European countries.

## 4. WORKSHOP RESULTS

The following sections provide a summary of the main items discussed during the workshop sessions as derived from the presentations as well as from the collected feedbacks of the audience.

### 4.1 Long Term Data Preservation Status

The Long Term Data Preservation issue and the ongoing European initiative coordinated by ESA with the involvement of the GSCB were addressed and extensively discussed during the 1<sup>st</sup> LTDP workshop held at ESA/ESRIN in May 2008 aiming at:

- Collecting and exchanging information on LTDP policies and technical approaches.
- Collecting feed-back for a potential European LTDP common approach and strive deriving common recommendations for it.
- Assessing the impact (benefits, drawbacks) of the proposed European LTDP Common Guidelines to each archive owner.

The need to preserve, make accessible and exploitable all available EO space data for long term science and environmental monitoring applications but also to support the United Nations Framework Convention on Climate Change (UNFCCC), the Climate Change initiative, GMES and other programmes also in the future was confirmed. The preservation of EO space data was referred as a responsibility of the Space Agencies or data owners as they constitute a humankind asset but it was highlighted the huge increase in terms of data volume expected in the next years (around 5-10 times in Europe) and the challenge for data owners to coherently preserve and provide access to these data. A cooperative and harmonized collective approach on Long Term Data Preservation (LTDP) in Europe (i.e. a European EO LTDP Framework) is necessary to coordinate and optimize European efforts in the LTDP field and to ultimately result in the preservation of the complete European EO space data set for the benefit of all European countries and users and with a reduction of overall costs.

The Long Term Data Preservation Working Group was formed at the end of 2007 with representatives from ASI, CNES, CSA, DLR and ESA within GSCB. Its goal is to define and promote the LTDP Common Guidelines with the involvement of all the European EO space data and archive owners and also to increase awareness on LTDP. **European EO LTDP Common Guidelines** have been generated by the LTDP working group and consolidated to take into account the results of the debate with all European and Canadian EO data owners during the 1<sup>st</sup> Earth Observation LTDP workshop held in May 2008 where they were considered as first step towards the implementation of the European LTDP Framework by the European stakeholders. European EO LTDP Common Guidelines were Issued in December 2009 after the completion of a public review process, updated in September 2010 (Issue 1.1) and are available at the URL:

[http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines\\_Issue1.1.pdf](http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_Issue1.1.pdf).

I

### 4.2 FIRST study

LTDP/FIRST project is part of the ESA LTDP programme and is aimed at:

- a) Capturing, understanding and consolidating Earth Science user needs and requirements in terms of long term preservation of data.

- b) Identifying and consolidate the composition of the datasets necessary to support Earth Science needs.
- c) Identifying associated knowledge necessary to support research and application activities.
- d) Assessing current European infrastructure (e.g. archives) capabilities.
- e) Providing inputs for the evolution of the LTDP Common Guidelines and for the set-up of the European LTDP Framework.

User requirements were collected through a questionnaire and interviews with the involvement of international experts, GEO Communities of practices and User Interface Committee.

Considerations arisen from user survey results are:

1. Preservation is a common and widely recognized need to allow performing today and future Earth Science research and application activities.
2. Preserve all forever is the mantra for the users.
3. A user centered vision should be applied rather than a provider centered one.
4. Preservation and future re-use of data paves the way to further knowledge growth.
5. Availability and access to all data should be open and free.
6. Harmonization among different actors and providers should be a primary objective.
7. Availability and accessibility to all data, documents and information is still an issue.
8. Reliability of data to be used should be addressed.
9. Access to data and services should be simple and costless.
10. Security issues are matter of providers/owners and should minimize the effort on user side.
11. Time-series durations/windows needs are different for different applications and utilizations however the long-term preservation is necessary (for more than 15 years for half of the users).
12. Periodic reprocessing of data is a necessity.
13. Apart some real-time applications, general performances are perceived as good or acceptable, and are not considered as an actual issue.
14. There is no dissimilar utilization of different EO data/products (e.g. SAR, Opt, multispectral, Hyperspectral, others) being all of them used and necessary in different Earth Science domains and applications.

The workshop confirmed the overview of Earth Science user's needs and requirements both for Earth Science and EO specific data preservation, identified in the frame of the ESA FIRST project. It has also integrated additional user needs and feedbacks to consolidate the Earth Science users' community requirements.

Stating from the consideration that long-term preservation is a necessity, the discussion was focused on "what" exactly should be the target of preservation. Looking at documents, survey and statistics there are some points of focal attention. These are:

1. Preservation must address all elements necessary to have the complete knowledge of when the specific object was generated. This means including all the amount of information necessary to rebuild today something that was generated 20 (or more) years ago. (*scope*)

2. Preservation must address the quality information relevant to the element/product. This serves to qualify all the elements participating different datasets. Quality covers concepts applicable to uncertainties and known errors as well as documents too. (*trustability*)
3. Coherency among all components of a specific dataset helps in qualifying the entire dataset. (*coherence*)

The most used and required EO data (categories C1 to C5) are of type L0, L1 and L2 and a wide variety of In-situ (C8, C9) and aircraft/balloon generated (C6, C7). There are situations (oceanography) where mainly In-situ data (C6, C7) are used. Common duration of time series needed for the different applications is for more than 15-20 years (46% of the total of users). These data have been compared with ordering/production statistics of ESA/ESRIN and Canadian Space Agency (CSA) and confirmed (further analysis are planned with additional production statistics). Data between survey and owners statistics are coincident.

Secondary data (auxiliary, ancillary, documents, etc) statistics are in line with the ones of primary data. Time series/windows are a point of attention. Concentrations around 5 years, 10/15 years and longer than 20 years have been identified inside the same Earth Science domain and commonalities identified among different Earth Science domains. This could help in better understanding the actual needs and performances profiles in terms of data series.

### **4.3 Comments and recommendations**

During the workshop the participants have confirmed and further consolidated the LTDP needs identified in the FIRST study.

GEOSS would need a convincing answer on LTDP by 2015 regarding:

- Broad convergence.
- Demonstrable progress towards improved preservation of data.
- Materialisation of initial benefits from the approach selected.

And an integrated approach across GEOSS for:

- Engaging communities to seek broad consensus.
- Strong connection to cross-cutting tasks, e.g. QA4EO and Metadata Harmonisation.
- Implementation support: capacity building, tools, archiving resources.

The European Commission recognises the need to continue the efforts – in coordination with EU Member States and seeking global partnerships – to support e-Infrastructures for modern science. It is of strategic importance the development of a fully functional e-Infrastructure for Scientific Data that is flexible, efficient, cross-disciplinary and cross border.

QA4EO provides a practical means to facilitate harmonisation and the robust assignment of a Quality Indicator to data and derived products with the objective to maximize the applicability and interoperability of EO data/products. EO end-users need to access quality information (e.g. accuracy and precision) on the products they use and therefore traceability of all processing steps (i.e. from acquisition to delivery) should be maintained and made available (e.g. to allow error propagation, reprocessing, development of new products, multi sensor inter-calibration, long-term studies, etc).

The creation of a data and software library involving institutions from eight European Countries through the ERA-ORA Project (CESBIO, FR; DLR, GE; Gamma A.G., CH; Landbouwniversiteit Wageningen, NL; Università Tor Vergata, IT; Universitat de Valencia, SP; Université Catholique de Louvain, BE; University of Sheffield, UK) was presented. The Library is maintained by the Tor Vergata University EOLab team (beyond the formal deadline). The motivation is to put radar data into practical use as experimental data at the disposal of a single researcher are (always)

limited. Joining different sets of experimental data acquired in the past by different institutions yields an ensemble of measurements sufficiently representative of both system and natural variability.

The objective of Climate Change Initiative is to realize the full potential of the long-term global Earth Observation archives that ESA together with its Member states have established over the last thirty years, as a significant and timely contribution to the ECV databases required by UNFCCC. The CCI programme needs all Low Bit Rate Data (including Meris FR), needs to have an efficient access to all data that are under the responsibility of ESA (ERS-1, ERS-2, Envisat, Explorers and third party missions), that includes historical data (e.g. AVHRR Global Land Data Set from 1992 to 1998 of extreme importance for the CCI Terrestrial ECVs); this includes also all auxiliary data (e.g. orbit, time correlations, housekeeping data, Calibration and Validation data) in addition to Level 0 as well as higher level processed data. The CCI projects need to know which data is available in the ESA archive for the past 30 years:

- Level 0, Level 1, Level 2 and higher;
- All Auxiliary Data sets required processing this data;
- Housekeeping Data (we need to know the on-board configuration when re-processing level 0);
- Calibration and Validation and efficient Data reprocessing.

CCI will also lead to re-calibration requirements and therefore reprocessing from level 0 (e.g. AVHRR data). This leads to a reprocessing capability able to reprocess complete data sets every two years (at least).

Calibration and Validation are the result of Engineers and Scientists work and are subject to improvement. In the context of “Climate Politics” every step shall be traceable and reproducible. Level 1 and Level 2 products are not “evidences” as subject to discussion between experts (Calibration and Validation).

During the workshop the CCI gave the following recommendations:

- The **responsibility** of preserving any data is of the Agency/Entity that acquired or produced this data.
- It is **fundamental** to preserve level 0 data and the associated housekeeping data;
- Level 1, 2 and higher should also be preserved with **all auxiliary data** needed;
- **Complete data sets** (i.e. all data required for the production of a given ECV) shall be **efficiently** retrievable from the archive.

LTDP should provide:

- **Reprocessing capacity** to allow to re-produce any data set to take into account feedbacks from programmes using this data (e.g. the CCI Programme).
- Typically "**less than every two years** - but it must produce a **fully consistent reprocessed data set** consistent across all missions (ERS\_1, ERS\_2 ,ENV) from start to end".

Data archive centers have contributed to give a better understanding of the issues addressed in the workshop. In particular the CEDA case was presented. CEDA provides acquisition of third party data for easy access for UK academic user community (including Envisat, MetOp, UK Met Office, ECMWF, ...) and provides preservation & dissemination of UK environmental research community's output datasets (including satellite data, derived products i.e. level 2, level 3, field campaign data like aircraft, ground based, and model output data). CEDA showed the needs of data user community derived by their experience: different data levels for different applications should be preserved and the data version must be identifiable. Regarding how long preserving data, if it's the primary archive: forever...with point and click interfaces not scalable – need

scripting access (e.g. for large scale processing). Climate modeling community requires data in CF-netCDF format particularly for higher level products (e.g. L3 gridded data). It is needed to follow standards for metadata and archives at “file level” and to have browsing capabilities to identify and pick the data needed with associated character (e.g. quality, metrics, annotations...). The WDCC has showed another important point of attention: the publication of Environmental data. Aspects most critical to be considered by LTDP are: storage location (the volume of data can be huge), formats (can be various), exposition/registry (it is not sufficient to save data “somewhere” on the web), quality (not all data are qualified), stability, identifier.

Data preservation is a need also for services for emergency response. During the workshop the SAFER (**S**ervices and **A**pplications **F**or **E**mergency **R**esponse) case was showed. It has confirmed that historic/archive data is having an important role for pre-disaster/crisis area assessment: quick access to on-line - at least near-line - historic archives is a must for the success of any emergency mapping activity. The required processing levels range from level 1a to 2 orthorectified. SAFER showed that the availability of historic datasets of uniform quality from heterogeneous sources (different scene sizes, different orthorectification methods, different modes/channels) are still an issue.

As regard the centres of reception and archiving of data and products, DLR-DFD has confirmed the importance of knowing who will use the data and how in order to anticipate future science needs.

Another area that must be considered to understand the various needs is the one addressing and performing reanalysis. At the workshop this field was represented by ECMWF's activities in Weather Prediction and Reanalysis that has confirmed the FIRST's preliminary study results. ECMWF for its activities use large amounts of EO data (from satellites and other sources) and it needs EO data with well-understood uncertainty and error characterization. The past observations (for period from 15 to 40 years) may be re-processed using a modern forecasting/assimilation system. Climate-quality reanalysis has demanding requirements like low random and systematic errors, high stability and consistency.

Data preservation is also fundamental for L-SIS Research Group's activities as a benefit to service quality (self-describing data i.e. data with sufficient metadata, data independence) and interoperability perspective based on OGC standardization work.

Several specific application cases were also presented at the workshop. Some of them are described in the following sections.

#### **4.3.1 OCEANOGRAPHY CASE**

In reference to science inherent the Oceanography, at the workshop was presented the GKSS, Institute for Coastal Research, Remote Sensing Group case. For their activities it is necessary to preserve lowest-level data (level-0 at NASA, level-0 and level-1 ESA) with metadata: all instrument calibration data (pre-launch and on-orbit), all knowledge of instrument & spacecraft anomalies, all state-changing events, storing of processing codes with sufficient information such that the higher level data (L1-L4) can be reproduced are also needed. For their application is necessary the availability of processors for step from level 1to2 and from level 2to3 and the storage of older versions of processors. They would like available a rich catalogue of metadata, harmonised, but adjusted to the respective domain and mission, evolving over time and their easy access like: descriptions to allow checks (not only ATDB, but also DPM, detailed processing model), timeline of changes to instruments/processing and cyclic reports, but focussed on stored data (quality). Equally important for them is to keep data format consistent over time or, if not possible, to reformat them or provide a converter; authorisation to access the data, e.g. by open access (and priorities), or at least for the relevant user groups and technical accessibility of data, i.e. online access in reasonable time and with simple and state-of-the-art

protocols is also considered important as also the organisation of storage in a way to support time line access also for very long time lines over an area (spatio-temporal cubes); it was also proposed for consideration the option to bring the algorithms to the data instead of data transfer and optimisation of storage organisation for typical retrieval scenarios of the user community of the respective data.

### 4.3.2 CRYOSPHERE CASE

In reference to science inherent the Cryosphere, at the workshop was presented the IPY experience (METNO). It has highlighted as very important the catalogue interoperability, in term of machine interface, standards (CEOS WGISS) and ensured linkages; availability of online data access, filtering in term of standardized vocabularies and interdisciplinary support where also highlighted as a need.

### 4.3.3 GEOPHYSICS CASE

In reference to science inherent the Geophysics, at the workshop was presented the NORMAP experience (consortium: NERSC, met.no, KSAT with international support from CERSAT). NORMAP showed some factors which in practice limit use of satellite data and the importance of someone who is part of the chain between satellite agencies and researches. Actually many researchers know a lot of geophysics but lack knowledge about satellite terminology, algorithms and resources to download and process large amounts of data. NORMAP can provide geophysical products in standardized simple-to-use formats, generate geophysical products with dedicated regional algorithms and time-averaged products. It can collect time series of various satellite products for certain points (e.g. positions of buoys or other in situ measurements), for certain small regions of interest, at high spatial resolution (e.g. a fjord or bay), co-locate products from various satellite sensors, e.g. to assist interpretation or for generation of multi-sensor products. For their activities NORMAP need as input any satellite data which can be used to retrieve geophysical parameters useful for marine and climate research, as long time series of level 1 and level 2 satellite data, ideally homogeneity over time and possibility an infrastructure to order satellite data in large amounts for selected regions.

## 4.4 Conclusions and Next Steps

The workshop has shared and provided an overview of Earth Science users needs and requirements for the data preservation captured in the frame of the FIRST project. It has also integrated additional user needs and feedbacks to consolidate the Earth Science users' community requirements providing a forum for Earth Science users to present their needs and expectations in terms of long term data preservation and access. Earth Science needs:

- **Interoperability and sharing** of existing infrastructures for the preservation and access of Earth Science data.
- **Sustainability** of the data infrastructure through the availability of funding mechanisms that enables all to contribute as well as to benefit, through an increased return on investment.
- Availability of **common policies** and approaches for the preservation and access to the data.
- The solution requires joining forces: **An international framework for collaboration and cooperation.**

The workshop demonstrated a continuous and growing interest on the subject of long-term data preservation and of the ongoing LTDP activities.

Several presentations were held covering different application domains and international initiatives.

The major points highlighted during the workshop are:

- Long-term preservation of Earth Science data is necessary. Preserve all forever is the mantra for the users.
- All elements necessary to have the complete knowledge on a specific data object need to be preserved in addition to the data itself.
- Availability and accessibility to all data, documents and information is still an issue.
- Quality information relevant to the element/product need to be generated and preserved together with the data.
- Coherency among all components of a specific dataset need to be ensured.
- Preservation of past data paves the way to further knowledge growth.
- Harmonization among different actors and providers should be a primary objective (for example in data access).
- Time-series durations/windows needs are different for different applications and utilizations however the long-term preservation is necessary (for more than 15 years for half of the users).
- Periodic and efficient reprocessing of data is a necessity.
- All Earth Science data categories (e.g. EO, In-situ, airborne and balloon) play an important role for scientific use.
- Permanent sources of reliable and ready to use data are needed.
- Improvement is needed in the access to historical datasets: the temporal – spatial cube.
- Availability of coherent set of historical data is needed.
- Commonly agreed vocabulary and taxonomy are needed.

Main comments and recommendations from the workshop participants were confirming the appreciation for the LTDP initiative coordinated by ESA and the recognition of its benefits for the different EO communities. A general support for the prosecution of the activities was expressed by the audience and the enlargement to the wider Earth Science domain was recommended.

Next steps can be summarised as follow:

### **FIRST Project**

1. Consolidation and further refinement of User Requirements exploiting the results of the workshop.
2. Identification and estimation of European/Canadian long-term data preservation infrastructure capabilities. The reference is constituted by the European LTDP Guidelines and a capability and maturity model will be used to estimate compliances and differences.
3. Proposal for updating of the European LTDP Guidelines according to study results.
4. Provision of recommendations for the way forward for the set-up of the European LTDP Earth Science Framework.

## 5. ANNEX 1 – WORKSHOP AGENDA



Workshop  
 Long term data preservation user needs  
 ESA ESRIN, 21 October 2010

*“Valorise the past for future needs”*

### Workshop Agenda (version 1.0 11/12/2010)

9:00	Registration	
	<b>Session 1 – LTDP and FIRST activities and results</b>	<b>Chair: Henri Laur</b>
9:30	<b>ESA</b> Welcome and Introduction	Henri Laur
9:50	European <b>LTDP</b> activities and guidelines	Mirko Albani
10:10	LTDP Activities and needs in <b>GEO</b>	Rob Koopman
10:30	<b>QA4EO</b> , preservation requirements for quality assurance	Giuseppe Ottavianelli
10:50	<b>EC</b> and European Scientific Data e-Infrastructures: an overview of the latest developments and future perspectives	Carlos Morais Pires
11:10	Coffee break	
11:30	<b>FIRST</b> Project preliminary results	Raffaele Guarino
	<b>Session 2 – LTDP User Needs</b>	<b>Chair: Mirko Albani</b>
12:00	An EC research project: ERA-ORA	Domenico Solimini
12:15	Needs from science: an <b>Oceanography</b> case	Hajo Krasemann
12:30	Needs from science: a <b>Cal/Val - Radiometry</b> case	Lydwine Gross-Colzy
12:45	<b>Climate Change Initiative</b> case	Pascal Lecomte

13:00	Lunch	
14:00	Environmental Science Information Requirements	Victoria Bennett
14:15	<b>GMES</b> Services case: SAFER	Hans Weber
14:30	Needs from science: the <b>DLR</b> case and experience	Katrin Molch
14:45	<b>ECMWF</b> case and experience	David Tan
15:00	The New OGC Coverage Model and its Services: Case and Perspectives	Peter Baumann
15:15	Needs from science: a <b>Cryosphere</b> case, IPY	Oystein Godoy
15:30	Needs from science: a <b>Geophysics</b> case	Knut-Frode Dagestad
15:45	Coffee break	
16:10	<b>DFG</b> (German Research Foundation) case and experience	Andreas Hense
16:30	<b>Evolution of ESA</b> services	Simon Jutz
	<b>Session 3 - Wrap-up and Discussion</b>	<b>Chair: Eugenia Forcada</b>
16:50	LTDP <b>Future vision</b>	Eugenia Forcada
17:10	Open discussion	Moderator
17:45	Conclusion	Mirko Albani / Raffaele Guarino
18:20	Social event - drink	

## 6. ANNEX 2 – LIST OF PARTICIPANTS

Name	Company	Country
Albani Mirko	ESA	EUROPE
Albani Sergio	ACS	ITALY
Alm Lars	Swedish Space Corporation (SSC)	SWEDEN
Aulamo Osmo	Finnish Meteorological Institute	FINLAND
Badiali, Lucio	INGV	ITALY
Baumann Peter	Jacobs University	GERMANY
Bennett Victoria	STFC, NERC Earth Observation Data Centre	UNITED KINGDOM
Ciguerre Christine	CSA	CANADA
Conway Esther	STFC	UNITED KINGDOM
Cornier Christine	CNES	FRANCE
Di Iorio Alessio	ROVSING A/S	DENMARK
Duplaa Michel	CNES	FRANCE
Fehr, Thorsten	ESA	ITALY
Forcada Eugenia	ESA	EUROPE
Förster Saskia	GFZ German Research Centre For Geosciences	GERMANY
Frithjof Barner	Euromap GmbH	GERMANY
Guarino Raffele	Capgemini IT	ITALY
Guidetti Veronica	ESA	ITALY
Hazane M.Philippe	CNES	FRANCE
Hense Andreas	DFG (German Research Foundation)	GERMANY
Hubbard Steven	Infoterra Ltd.	UNITED KINGDOM
Jutz Simon	ESA	EUROPE
Koopman Robert	GEO	SWITZERLAND
Krasemann Hajo	GKSS Research Centre Institute for Coastal Research	GERMANY
Landgraf Guenther	ESA	EUROPE
Larroque Martine	CNES	FRANCE
Laura Henri	ESA	ITALY
Lavender Samantha	ARGANS Limited, Unit 3	UNITED KINGDOM
Lecomte Pascal	ESA CCI	EUROPE
Leone Rosemarie	ESA	EUROPE
Marchetti Pier Giorgio	ESA	ITALY
Mason Graeme	ESA	EUROPE
Mikush Eberhard	German Aerospace Center - DLR	GERMANY
Molch Katrin	German Aerospace Center - DLR	GERMANY
Monsorno Roberto	EURAC Research	ITALY
Morais-Pires Carlos	European Commission	EUROPE
Moreno Richard	CNES	FRANCE
Nityaporn Sirikan	ESA	EUROPE
Ottavianelli Giuseppe	ESA	EUROPE
Øystein Godøy	Meteorological institute	NORWAY
Padovano Michele	Capgemini	ITALY
Pérez Moreno Rubén F.	GMV	SPAIN
Perez Navarro Oscar	GMV	SPAIN
Pinnock Simon	ESA	EUROPE
Polsinelli Barbara	Capgemini IT	ITALY

Rempicci Luciano	Capgemini	ITALY
Solimini Domenico	Tor Vergata University	ITALY
Tan David	ECMWF	UNITED KINGDOM
Vahid Naeimi	DLR-DFD	GERMANY
Van Hees Richard	SRON Netherlands Institute for Space Research	NETHERLANDS
Weber Hans	DLR	GERMANY
Yang Jingsong	ESA trainee	EUROPE
Zhang Lu	ESA Trainee	EUROPE

Note: some registered people were not able to attend the workshop due to problems with flights.