



DOCUMENT

Long Term Data Preservation Functional User Requirements Document LTDP/FURD

Prepared by M. Albani (ESA), R. Guarino (Cap Gemini), R. Leone (ESA)
Reference LTDP-GSEG-EOPG-RD-11-0004
Issue 3
Revision 0
Date of Issue 04.05.2011
Status Under Review
Document Type User Requirements
Distribution



CHANGE LOG

Reason for change	Issue	Revision	Date
Move to ESA standard document	2	0	06.04.2011
Implement Comment	3	0	04.05.2011

CHANGE RECORD

Issue 1	Revision 0		
Reason for change	Date	Pages	Paragraph(s)
Move to ESA standard document	06.04.2011	all	all
Implement Comment	04.05.2011	all	all

Table of contents:

1	INTRODUCTION.....	6
1.1	LTDP context.....	6
1.2	Document scope	6
1.3	Document structure.....	8
1.4	Applicable and reference documents.....	8
1.4.1	Applicable.....	8
1.4.2	Reference.....	9
1.4.3	References for User Requirements definition	10
1.4.4	References for Cross Analysis.....	12
1.4.5	Standards	14
1.5	Acronyms and Definitions	14
1.6	Data categories and classification.....	15
1.6.1	Documentation List	16
1.6.2	Earth Observation Space Data	17
1.6.3	Airborne and Balloon Data	18
1.6.4	In-situ Data.....	19
1.6.5	Earth Observation sensors Sub-categories classification	20
1.7	List of Earth Science fields and applications domains considered	22
1.8	List of international initiatives and programmes analysed	26
1.9	Campaigns	26
2.	REQUIREMENTS.....	29
2.1	Requirements Classification	29
2.2	Survey Population Statistics	30
2.3	Common requirements	32
2.4	Atmosphere Domain	34
2.4.1	Domain Description	34
2.4.2	General data needs	35
2.4.3	Earth Science Data User Requirements	36
2.4.4	EO Data User Requirements	37
2.4.5	Additional Data User requirements.....	38
2.4.6	Summary Earth Science Dataset Composition.....	38
2.4.7	Special cases.....	39
2.5	Biosphere.....	40
2.5.1	Domain Description	40
2.5.2	General Data Needs.....	42
2.5.3	Earth Science Data User Requirements	42
2.5.4	EO Data User Requirements	44
2.5.5	Additional Data User requirements.....	44
2.5.6	Summary Earth Science Dataset Composition.....	45
2.5.7	Special case.....	46
2.6	Glaciology (Cryosphere)	47
2.6.1	Domain Description	47
2.6.2	General Data Needs.....	48
2.6.3	Earth Science Data User Requirements	48
2.6.4	EO Data User Requirements	49
2.6.5	Additional Data User requirements.....	50
2.6.6	Summary Earth Science Dataset Composition.....	51
2.6.7	Special case.....	52
2.7	Oceanography.....	53
2.7.1	Domain Description	53
2.7.2	General Data Needs.....	53



2.7.3	Earth Science Data User Requirements	54
2.7.4	EO Data User Requirements	55
2.7.5	Additional Data User requirements.....	57
2.7.6	Summary Earth Science Dataset Composition.....	57
2.7.7	Special cases.....	57
2.8	Geophysics.....	58
2.8.1	Domain Description	58
2.8.2	General Data Needs.....	59
2.8.3	Earth Science Data User Requirements	59
2.8.4	EO Data User Requirements	60
2.8.5	Additional User data requirements	61
2.8.6	Summary Earth Science Dataset Composition.....	62
2.8.7	Special case.....	62
2.9	Earth as a system (EaaS).....	63
2.9.1	Domain Description	63
2.9.2	General Data Needs.....	64
2.9.3	Earth Science Data User Requirements	64
2.9.4	EO Data User Requirements	65
2.9.5	Additional User requirements.....	66
2.9.6	Summary Earth Science Dataset Composition.....	66
2.9.7	Special case.....	67
2.10	Cal/Val	68
2.10.1	Domain Description	68
2.10.2	Cal/Val Engineering Data Needs	69
2.10.3	Earth Science Data Requirements from Cal/Val Engineering.....	69
2.10.4	Earth Science Dataset Composition.....	71
2.10.5	Additional requirements.....	71
2.10.6	Special case.....	71
2.11	Geoinformatics	72
2.11.1	Domain Description	72
2.11.2	General Data Needs.....	72
2.11.3	Earth Science Data User Requirements	72
2.11.4	EO Data User Requirements	73
2.11.5	Additional User requirements.....	74
2.11.6	Summary Earth Science Dataset Composition.....	74
2.11.7	Special case.....	74
2.12	GMES	75
2.12.1	General Data Needs.....	75
2.12.2	Earth Science Data User Requirements from GMES Services	76
2.12.3	Example: GMES Services Dataset Composition	76
2.12.4	EO Data Requirements from GMES Services	77
2.12.5	Additional requirements.....	78
2.12.6	Earth Science Dataset Composition.....	78
2.12.7	Special case.....	79
2.13	Climate Change Initiative (CCI).....	80
2.13.1	Climate Change Initiative Data Needs.....	80
3.	SUMMARY OF REQUIREMENTS AND NEEDS	81
3.1	Summary of needs and requirements from ES user community.....	82
4.	SUMMARY OF EARTH SCIENCE DATA CATEGORIES AND COMPOSITION OF THE RELATED DATASET TO BE PRESERVED	84
4.1	Earth Observation (space-born) data needs	84
4.2	EO dataset composition.....	84
4.3	Earth Science data needs.....	87
4.4	Earth Science dataset composition.....	88





1 INTRODUCTION

1.1 LTDP context

The long term preservation of Earth Science 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 programme. ESA in this context is running for a period of three years a Long Term Data Preservation Programme targeted to Earth Observation space data. ESA is also coordinating an initiative for the set-up of a European EO Long Term Data Preservation (LTDP) Framework, with the involvement of major European Earth Observation stakeholders, and aimed at guaranteeing the preservation and access to European Earth Observation space data with possible enlargement to Earth Science data.

Long-term accessibility and exploitability of Earth Science data requires that not only data but also the associated knowledge (e.g. technical and scientific documentation, algorithms, data handling procedures, etc) needed to guarantee their understandability is properly preserved and made accessible. Earth Science is a very challenging domain as it is inherently very broad and scattered with as many as nine main categories of data. Within each category completely different source instruments are available, generally owned and operated by different entities applying different data preservation policies – or none at all. Interoperability, application of common and standard preservation approaches and policies and the utilization of harmonised technological solutions and services in this context are key aspects to guarantee that ES data will be available and understandable in future to allow current and next generations of scientists to exploit them for very sensitive applications.

This document is one of the deliverables produced in the Long Term Data Preservation (LTDP) FIRST Study carried out as part of the ESA LTDP programme. The study is conducted with the support of Ground Segment Coordination Body (GSCB) partners as part of the coordination activities ongoing in Europe under ESA lead for the set-up of the European EO LTDP framework.

1.2 Document scope

The LTDP/FIRST study is aimed, among other activities, at capturing and understanding Earth Science users' needs in terms of long term preservation (including accessibility and exploitability aspects) of Earth Science data and associated knowledge necessary to support their research and application activities. The study considers necessities and expectations from users pertaining to all the Earth Science domains and from international initiatives like the United Nations Framework Convention for Climate Change and addresses all possible Earth Science data categories including, but not limited to, Earth Observation space data. FIRST is therefore to be considered as an important step for the extension of the current European LTDP cooperation framework from the Earth Observation domain to the wider Earth Science concept.



This document is aimed at describing User Requirements and needs in terms of long-term preservation (including accessibility and exploitability aspects) of Earth Science data. Requirements are expressed mainly considering Earth Science (ES) necessities in terms of “*what*” is necessary to preserve and “*why*”. In other words, the document focuses on the “*user business view*”. Other dimensions (i.e. *when, how*) are considered out of scope.

The users considered in this study are from all the communities with a vested interest in Earth science and caring of scientific (i.e. scientists and researchers) or application aspects (e.g. value adders or GMES services) and users like decision makers (e.g. civil protections) or international initiatives (e.g. Climate Change initiative).

The “*Who*” dimension is on the other hand based on the concept that the European LTDP framework is a collaborative environment among many stakeholders that should provide their contribution to guarantee that European Earth Science data and associated knowledge are properly preserved and made accessible and exploitable to users.

Hence, the scope of this document is to address all possible Earth Science domains with the relevant fields of application (ref. 1.7 List of Earth Science fields and applications domains considered).

Furthermore, analysis of international initiatives and activities that might have needs in terms of Earth Science data has been performed in order to gather additional requirements.

The geographical scope of this document privileges European and Canadian users and communities (e.g. CoPs) however requirements collection and analysis have also been conducted in non-European countries (e.g. United States entities like NASA, USGS, NOAA, Library of Congress/Columbia University, etc; see Table 4 - References for cross comparison).

The document is 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 Earth Science domain.

This document will not focus on architectures and/or technologies to be used for the purpose that are objective of other projects.

This document collects Earth Science (ES) requirements and proves a more detailed focus on Earth Observation (EO) satellite sensed data requirements. Aspects like the need of long-term time series of data and preservation of data for long time are also considered with particular attention.

Requirements have been collected by:

- Participating to events, seminars and workshops having relation with Earth Science data requirements and utilization (e.g. LIMES, Living Planet Symposium, etc.). The focus was to understand and register what kind of data and information are used (or would be necessary to use) for the purpose of scientist, service providers and value adders.



- Performing a survey through analysis of questionnaires (one generic and one specific). Questionnaires were sent to scientific communities of practice, public organizations (e.g. Universities and Research Centers, civil protections) and value adders (e.g. commercial entities).
- Interviewing researchers and investigators of different fields (e.g. forestry, oceanography, atmosphere, climatology, etc.).
- Analyzing the needs and expectations of different communities to identify commonalities (e.g. I/Rinascimento Digitale, NASA/Planetary Data System, USA/Library of Congress, InterPARES, CERN, and others; see Table 4 - References for cross comparison).
- Analyzing existing legislation, regulations or rules of organizations having as an objective long-term data preservation or retention (e.g. OECD rules, ICSU/WGCC, ARIADNE, INSPIRE, others).

This document will be kept under a change management process for any possible future evolution.

1.3 Document structure

The document consists of four different chapters:

The first chapter contains the introduction and the scope of the document plus all the categories, classifications and references necessary to support requirement definitions. This chapter contains all hyperlinks useful to access directly to used data and information.

The second chapter deals with user requirements and contains a short introduction to each Earth Science domain, a general description of needs for each science domain and tables of user requirements relevant the domain and relevant the specific EO dataset. This chapter includes the two specific cases of GMES and CCI.

Chapter 3 provides a short summary and some points of attention.

Chapter 4 addresses the Earth Science dataset composition summary.

1.4 Applicable and reference documents

1.4.1 Applicable

Nr	Type	Description
AD1	FIRST SoW	LTDP-GSEG-EOPG-SW-09-0004, issue 1.0, dated



		6/11/2009
--	--	-----------

Table 1 - Applicable documents

1.4.2 Reference

Nr	Type	Description
RD1	LTDP Guidelines	Long Term Preservation of Earth Observation Space Data – European LTDP Common Guidelines, version 1.1 30/09/2010
RD2	LTDP EO datasets	Long Term Preservation of Earth Observation Space Data – European EO Space data set, version 1.0 12/08/2010

Table 2 - Reference documents



1.4.3 References for User Requirements definition

The following Table 2 shows the documents used to identify and capture User Requirements as in the scope of this document.

The ID defines the Source of Requirement (**SOR**) with a unique number for complete identification. This ID is referenced in requirement’s description in the column **Source**. Hence it allows traceability of sources used to build requirements.

Those documents and references are considered good source of requirements because:

- a) Are the formal ones published by European organisations having responsibilities in scientific research and cooperation.
- b) Constitute the outcome of activities and works of scientific researchers (single person, communities of practices, entities and research centres).
- c) Represent or include needs or expectations in terms of data utilisation and preservation (e.g. SOR-1 clearly address preservation needs, Action C17) issued by European organisation or international organisations participated by European representatives.
- d) Some sources (e.g. GMES projects/services) are worth for LTDP/FIRST being indirect demanders of preservation (e.g. atmospheric circulation models applied by GMES) for their services.

ID	Description	Entity or Program
SOR-1	http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf	GCOS
SOR-2	Global Monitoring for Environment and Security programme (GMES) Services: <ul style="list-style-type: none"> a) Geoland2: g2_EL-RP-D_EL010-1_UserRequirements1100.pdf b) MyOcean:http://www.gmes.info/pages-principales/projects/myocean/ c) SAFER: http://www.emergencyresponse.eu/site/FO/scripts/myFO_contenu.php?noeu_id=32&lang=EN d) MACC, URL: www.gmes-atmosphere.eu/documents/.../D_D-INSITU_1.2_part2.pdf (document D_D-INSITU_1.2.pdf in-situ data) , D-GRG3_4°_MACC_CRG_BC_services_specs.pdf 	GMES



	e) G-Mosaic: http://www.gmes-gmosaic.eu/ (html page)	
SOR-3	<p>Climate Change Initiative</p> <ul style="list-style-type: none"> Climate Modelling User Group, Deliverable 1.2, Requirement Baseline Document (doc. CMUG D1 2_URD v 1 2.pdf to be required at http://www.cci-cmug.org/) CCI EO Data requirements. Mr. Peter Regner (ESA) & CCI Projects Data Experts (<i>draft version under approval process</i>) D1.1 Profile and needs of CMC.pdf, CMUG References to the Essential Climate Variables (ECV) concepts gcos-107.pdf (available at www.wmo.int/pages/prog/gcos/Publications/ and at http://www.wmo.int/pages/prog/gcos/Publications/gcos-143.pdf) 	CCI
SOR-4	<p>European Strategy Forum on Research Infrastructure</p> <ul style="list-style-type: none"> esfri_roadmap_update_2008.pdf (available at ftp://ftp.cordis.europa.eu/.../esfri_roadmap_update_2008.pdf) CESSDA_PPP_Projects.pdf (available at www.cessda.org/project/doc/CESSDA_PPP_Projects.pdf) esfri_e_irg_report_data_management:December_2009_en.pdf (available at http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=publications, and search the document) 	ESFRI
SOR-5	<p>COMMISSION REGULATION (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council, LexUriServ.pdf (data & metadata)</p> <p>COMMISSION REGULATION (EU) No 268/2010 of 29 March 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards the access to spatial data sets and services of the Member States by Community institutions and bodies under harmonised conditions, (access and usability)</p>	Inspire JRC
SOR-6	<p>GCOS Essential Climate Variable (ECV) and Fundamental Climate Data Record (FCDR)</p> <p>http://gosis.org/ios/MATRICES/ECV/ECV-matrix.htm</p> <p>http://www.wmo.int/pages/prog/gcos/Publications/gcos-143.pdf</p>	GCOS
SOR-7	<p>United Nations Framework Convention on Climate Change (ref. to Articles 4 and 12 of the Climate Change Convention)</p> <p>Subsidiary Body for Scientific and Technological Advice December 2003, and GCOS Climate Monitoring Principles report (ref.</p>	UNFCCC



	http://unfccc.int/)	
SOR-8	Group on Earth Observation and The Global Earth Observation System of Systems (GOSS) URL: http://www.earthobservations.org/ Starting reference document: http://www.earthobservations.org/documents/work%20plan/geo_wp0911_rev2_091210.pdf	GEO

Table 3 - Reference documents for requirements

1.4.4 References for Cross Analysis

The following table shows the document used to analyze similar activities or having similar long-term preservation objectives being still in progress or recently completed.

The ID defines the Reference To Similar (RTS) with a unique number for complete identification. This ID is referenced when of interest along the document. These additional initiatives have been analysed in the study and used to:

- a) Compare Earth Science LTDP needs and requirements with the ones from similar initiatives or initiatives having some affinities.
- b) Capture generic guidelines and best practices.
- c) Have better understating of rationales and capture lessons from others even if in different fields.
- d) Capture information to build the questionnaire and survey

These references do not directly generate requirements but have been used to improve and complete the requirements generated in the study. These are therefore traced in the document as RTS-XX.

ID	Description	Entity or Program
RTS-1	Global Observing Systems Information Center, <ul style="list-style-type: none"> • http://gosis.org/ (list of additional GOSIC internal references) • http://gosis.org/publications/pub_search.asp (details on ECVs) • http://gosis.org/gcos/GSN-flow.htm (surface stations network) • http://gosis.org/gcos/guan-flow.htm (upper air data) • http://gosis.org/goos/GTSP-data-flow.htm (salinity) • gcos-60.pdf (general data management requirements) 	GOSIC
RTS-2	National Oceanic and Atmospheric Administration's and National Climatic Data Center <ul style="list-style-type: none"> • sds_info.pdf (scientific data stewardship) 	NOAA NCDC

	<ul style="list-style-type: none"> • 037.pdf (MODIS specific project) • Environmental data management at NOAA. Archiving, stewardship and access, The National Academic Press, Paperback • 116809.pdf (files available at http://www.climate.noaa.gov)	
RTS-3	GAO-07-1172 – Climate Change Research document (available at http://www.gao.gov/new.items/d071172.pdf)	GAO
RTS-4	USA, The National Archives and Records Administration (NARA) Federal regulations <ul style="list-style-type: none"> • CFR 1228 - Disposition of Federal Records • CFR 1234 - Electronic Records Management 	USA NARA
RTS-5	http://gsa.confex.com/gsa/2010AM/finalprogram/abstract_18255_1.htm	GSA
RTS-6	2006 DataPreservation.pdf (long term plan) http://datapreservation.usgs.gov/resources.shtml	USGS
RTS-7	High Energy Physics (HEP) Data Preservation working group South.pdf DPHEP_ICFA_2009_public_02.pdf (available at CERN documentation service)	CERN HEP
RTS-8	Digital Preservation Europe (DPE) dp_in_high_energy_physics.pdf preservation-of-earth-system-model-data.pdf (available at http://www.digitalpreservationeurope.eu/publications)	DPE
RTS-9	United States of America (USA), North Carolina Geospatial Data Archiving Project (NCGDAP), URL: http://www.geomapp.net/docs/geomapp_survey_state.pdf	NCGDAP
RTS-10	CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval – European Union project. URL: http://www.casparpreserves.eu/	CASPAR
RTS-11	CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval – European Union project. URL: http://www.casparpreserves.eu/	CASPAR
RTS-12	http://dels.nas.edu/Study-In-Progress/Assessing-Requirements-Sustained-Ocean-Color/DELS-OSB-09-01 https://darchive.mblwhoilibrary.org/bitstream/handle/1912/2709/BakerChandler_finalDraft.pdf?sequence=1 http://journals.eecs.qub.ac.uk/codata/journal/contents/4_05/4_05_pdfs/DS401.pdf	NAS (Oceanography)
RTS-13	IGBP www.igbp.net	IGBP (Atmosphere)
RTS-14	International Glaciological Society (IGS) , http://www.igsoc.org/ International Association of Cryospheric Sciences (IACS), http://www.cryosphericciences.org/ International Permafrost Association (IPA), http://ipa.arcticportal.org/ NASA, http://ice.nasa.gov/	Glaciology / Cryosphere



RTS-15	The Global Sea Level Observing System http://www.gloss-sealevel.org/ ; ENSO, http://www.esrl.noaa.gov/psd/enso/	Oceanography/ Hydrology
RTS-16	NOAA, http://nomads.ncdc.noaa.gov/ Note: The NOAA National Operational Model Archive and Distribution System (NOMADS) is a Web-services based project providing both real-time and retrospective format independent access to climate and weather model data. Some early marine data starts on 1662 A.D.	Climate and Weather
RTS-17	Geophysical data web sites list data_portals_holders_websites.pdf URL : http://www.un.org/Depts/los/clcs_new/sources/	Geophysical data
RTS-18	NOAA, GEODAS http://www.ngdc.noaa.gov/mgg/geodas/geodas.html	Geophysical data

Table 4 - References for cross comparison

1.4.5 Standards

ID	Description	Short form
STD-1	ECSS-M-ST-40 C Rev.1	ESA ECSS
STD-2	Open Archival Information System http://public.ccsds.org/publications/archive/650x0b1.pdf	OAIS

Table 5 - Reference to standards

1.5 Acronyms and Definitions

CAL/VAL	<p>Calibration and Validation (CEOS definitions)</p> <p>Calibration is the process of quantitatively defining the system responses to known, controlled signal inputs.</p> <p>The process of assessing, by independent means, the quality of the data products derived from the system outputs</p>
CoP	Community of Practice, identifies a number of people having a



	common interest
ECV	Essential Climate Variables, ref. ESA Climate Change Initiative (CCI)
EO	Earth Observation, relevant to remote sensed imagery, physical and chemical parameters of the Earth.
FCDR	An FCDR denotes a long-term data record, involving a series of instruments, with potentially changing measurement approaches, but with overlaps and calibrations sufficient to allow the generation of homogeneous products providing a measure of the intended variable that is accurate and stable enough for climate monitoring. FCDRs include the ancillary data used to calibrate them. Definition from SOR-1 GCOS-138.pdf.
LTDP	Long Term Data Preservation
SAR	Synthetic Aperture Radar is a form of radar in which sophisticated post-processing of radar data is used to produce a very narrow effective beam.
SOR-xx	Source Of Requirement , within this document identifies the source of one or more requirements. It generically indicates a document or a program/project
VOC	<p>Volatile Organic Compounds (VOC); European Union definition and reference is:</p> <p>A VOC is any organic compound having an initial boiling point less than or equal to 250 °C measured at a standard atmospheric pressure of 101.3 kPa and can do damage to visual or audible senses. Ref. ^ "Directive 2004/42/CE of the European Parliament and of the Council" EUR-Lex, European Union Publications.</p>

Table 6 - Reference to standards

Definitions reported in Annex 4 of [RD-1] are considered applicable. In particular it is adopted the definition of Ancillary and Auxiliary terms.

1.6 Data categories and classification

Data categories and classifications used in this document are derived from the ones reported in the [Reference document RD-1].



Some extensions and changes have been introduced to improve taxonomy and those extensions are suggested to be used for amendment of the above document as well.

Data set have been divided into the following main categories:

1. **C1:** SAR imaging missions/sensors, high and very high resolution (different radar bands).
2. **C2:** Multi-spectral imaging missions/sensors, high and very high resolution.
3. **C3:** Medium resolution Land and Ocean monitoring missions/sensors (e.g. wide swath ocean colour and surface temperature sensors, altimeter, etc).
4. **C4:** Atmospheric chemistry missions/sensors.
5. **C5:** Other Scientific missions/sensors.
6. **C6:** Airborne (e.g. digital cameras single/multiple, digital line scanners, radar, laser topographic/bathymetric, etc). Helicopter Observation Platforms (HOPs) are considered in this category.
7. **C7:** Balloon (e.g. geomagnetic instruments, wind, temperature, radiation, radio propagation, particles, optical properties, chemistry, etc).
8. **C8:** Ground (e.g. seismography, temperature, humidity, wind, pressure, radiation, radiance, pollution factors, rain, chromatography, soil property, etc).
9. **C9:** Hydro (e.g. temperature, salinity, pollution factors, wind, pressure, water flow/flux/level, etc). Here are included data coming from buoys as well as from ships or other means to capture local data.

The first five categories (C1 to C5) can be grouped and referred as “Earth Observation Space data”, categories C6 and C7 can be grouped and referred as “Airborne and Balloon” data, whilst the remaining two categories (C8 and C9) can be grouped and referred as “In-Situ data”.

1.6.1 Documentation List

In the following paragraphs, a reference to **mission related documentation** is made. Here are considered documents relevant satellites, ships, buoys, planes, balloons platforms without distinctions. This section provides a general list of those documents:

The list of documents includes:

1. Mission architecture documents describing purpose, scope and performances of the mission and of the onboard instruments, information relevant orbits, platform position, attitude, ground coverage (acquisition footprint), head-roll-pitch.
2. Data/products descriptions, documents where raw telemetries and/or processed levels are described, here are considered data and products formats and characteristics too.
3. Documents describing measurement requirements and/or measurement performances (theoretical models). Documents drawing instruments characteristics, performances and instrument description (physical implementations). Documents describing models and/or algorithms needed (used) to obtain mission data and products including specific/special cases, known errors and configuration necessities. In other words, it should be provided documents covering conceptual environment, its implementation and its operations.



4. Reports concerned with measurement trends, failures, changes of performances, and out of service for any reason.
5. Reports and outcomes from events like congresses, studies, communities and investigators concerned with models' review, algorithm changes and Cal/Val changes affecting data processing chains
6. Documents related to the process of data qualification: precision, numerical representations, formats, uncertainties, errors, adjustment/correction methods (e.g. Cal/Val documents).

Campaign description documents where the following themes are addressed: objectives, scientific disciplines, geographic regions, data types and availability, results.

This list will be simply referenced as *Mission related documentation*.

1.6.2 Earth Observation Space Data

For C1, C2, C3, C4 and C5 categories, the composition of the data set to be preserved in the long term in order to allow exploitability and understandability by users consists of [RD-1]:

a) Primary Data:

1. Raw data (as acquired by the satellite and recorded at the stations or received via Third Parties) and/or Level 0 data.
2. Global or higher level mission products when systematically generated as part of the mission requirements and/or reprocessed.

b) Secondary data:

1. Ancillary data: (example list : Navigation and attitude data, Observation counts, Calibration temperature counts, Antenna temperature coefficients, Time (different measurements), Latitude, Longitude, Sun azimuth, Sun elevation, Earth incidence , Earth azimuth, Data quality). **See Note 1**.
2. Auxiliary data required to process the telemetry payload data to generate the nominal mission products. Generally auxiliary data contains parameters for correct algorithm processing. Example list: scaling errors, misalignment parameters, gridding/re-gridding tolerances, AOCS corrections, solar radiance, non-linearity corrections, temperature correction factors, mid-long term measurement channel drift.
3. CAL/VAL databases whenever available (including processing/reference validation data sets). This includes history of calibration/validation and the qualification process used and data reliability (e.g. uncertainty, errors).
4. *Mission related documentation*.

c) Metadata

- d) Browse images when generated

For each of the data listed in points a) and b) also the following information has to be generated and preserved: Representation Information (structure, semantic and other representation information), Packaging Information, Preservation Descriptive Information (Reference, Provenance, Context, Fixity and access conditions and rights). This information is needed to allow understandability and usability of the data from the users according to the Open Archival Information Standard (OAIS) (ref. Table 4 STD-2).

Note 1: Provided list of ancillary data is concerned with Earth Observation processing from raw data up to Level 2 maximum. It is possible to have different types of ancillary data for upper levels of processing. E.g. Level 4 Soil erosion assessment product could use as ancillary data rainfall reports and data in the observed period.

1.6.3 Airborne and Balloon Data

For C6 and C7 categories data composition is divided into:

- a) Primary Data:
- 1.Raw data (as acquired by the platform and recorded at the stations or received via other parties).
 - 2.Interpolated or processed data when systematically generated as part of the mission requirements and/or reprocessed.
- b) Secondary Data:
- 1.Ancillary data (e.g. georeference, GPS, DEM, etc.).
 - 2.Auxiliary data required to process the telemetry payload data to generate the nominal mission products (e.g. attitude, motion, ambient state, etc.).
 - 3.CAL/VAL databases whenever available (including processing/reference validation data sets). This includes history of calibration/validation and the qualification process used and data reliability (e.g. calibration methods, uncertainty/errors).
 - 4.Mission related documentation.
- c) Metadata (if any).
- d) Browse images when generated.

For each of the data listed in points a) and b) also the following information has to be generated and considered: Representation Information (structure, semantic and other representation information), Packaging Information, Preservation Descriptive Information (Reference, Provenance, Context, Fixity and access conditions and rights). This information is needed to allow understandability and usability of the data from the users according to the Open Archival Information Standard (OAIS) (ref. Table 4 STD-2).

1.6.4 In-situ Data

In-situ data are captured using an instrument in a specific position at a certain time. In-situ data belongs to one or both of the abovementioned categories C8 and C9. In literature and documents, in-situ data are generally mentioned by the type of sensor (or type of measure) assuming the specific location is known or described elsewhere. In-situ data can be a single datum or grouped by types, times, location or pre-transformed by algorithms.

For C8 and C9 categories data composition is divided into:

- a) Primary Data:
 1. Raw data (instantaneous non interpolated data as acquired by the sensor/platform and recorded at the stations or received via other parties).
 2. Interpolated data (by specification of short, mid and/or long term interpolation) as processed data when systematically generated as part of requirements and/or reprocessed.
- b) Secondary Data:
 1. Ancillary data (e.g. georeference, spatial resolution, positional drifting, GPS, DEM, weather conditions, etc.).
 2. Auxiliary data required to process the telemetry payload data to generate the nominal mission products (e.g. meteorological data, oxidation profile, etc.).
 3. CAL/VAL databases whenever available (including processing/reference validation data sets). This includes history of calibration/validation and the qualification process used and data reliability (e.g. calibration methods, uncertainty/errors), calibration drifting and reference levels.
 4. *Mission related documentation.*
- c) Metadata (if any)
- d) Browse images when generated

Interpolation periods, procedures and references are considered part of b.2) Auxiliary Data and b.4) Documentation.

For each of the data listed in points a) and b) also the following information has to be generated and considered: Representation Information (structure, semantic and other representation information), Packaging Information, Preservation Descriptive Information (Reference, Provenance, Context, and Fixity). This information is needed to allow understandability and usability of the data from the users according to the Open Archival Information Standard (OAIS) (ref. Table 4 STD-2)

The list of in-situ data/type captured from the questionnaires, interviews and other documents (ref. Table 4 - References for cross comparison) is provided below:

1. Temperature.
2. Wind Speed / Direction.
3. Wind Chill.
4. Humidity.
5. Solar Radiation.
6. Pollution factors (CO₂, NO₂...).
7. Rainfall/Rain gauge.
8. Pyrheliometer (or solar radiance).
9. Barometric Pressure (Air pressure).
10. Soil Moisture.
11. Soil Temperature.
12. Leaf Wetness.

13. Sap Flow (volume and speed).
14. Dendrometer.
15. Chromatographer.
16. Microphone (e.g. for earthquakes monitoring).
17. Seismometer.
18. Water flux/level (e.g. river monitoring).
19. Sea temperatures from buoys.
20. Salinity.
21. Ocean / Water current direction.
22. Surface reflectance.
23. Aerosol optical thickness.
24. Spectral ground reference & measure.

Here the list is presented **for information only** to support clarity of this document. This list should only be considered to be representative of data types and should not be regarded as complete and exhaustive.

1.6.5 Earth Observation sensors Sub-categories classification

Remotely sensed EO data included in the categories from C1 to C5 have been additionally sub-classified according to the following Table 7. The subcategories are needed to differentiate within the same category instruments of different type (e.g. radiometers and altimeters in C3) or instruments of the same type but with different characteristics (e.g. band, resolution).

Figure 1 shows the generic map used to classify sensors detailed in Table 7. Classification is coherent with mapping provided in [RD-2].

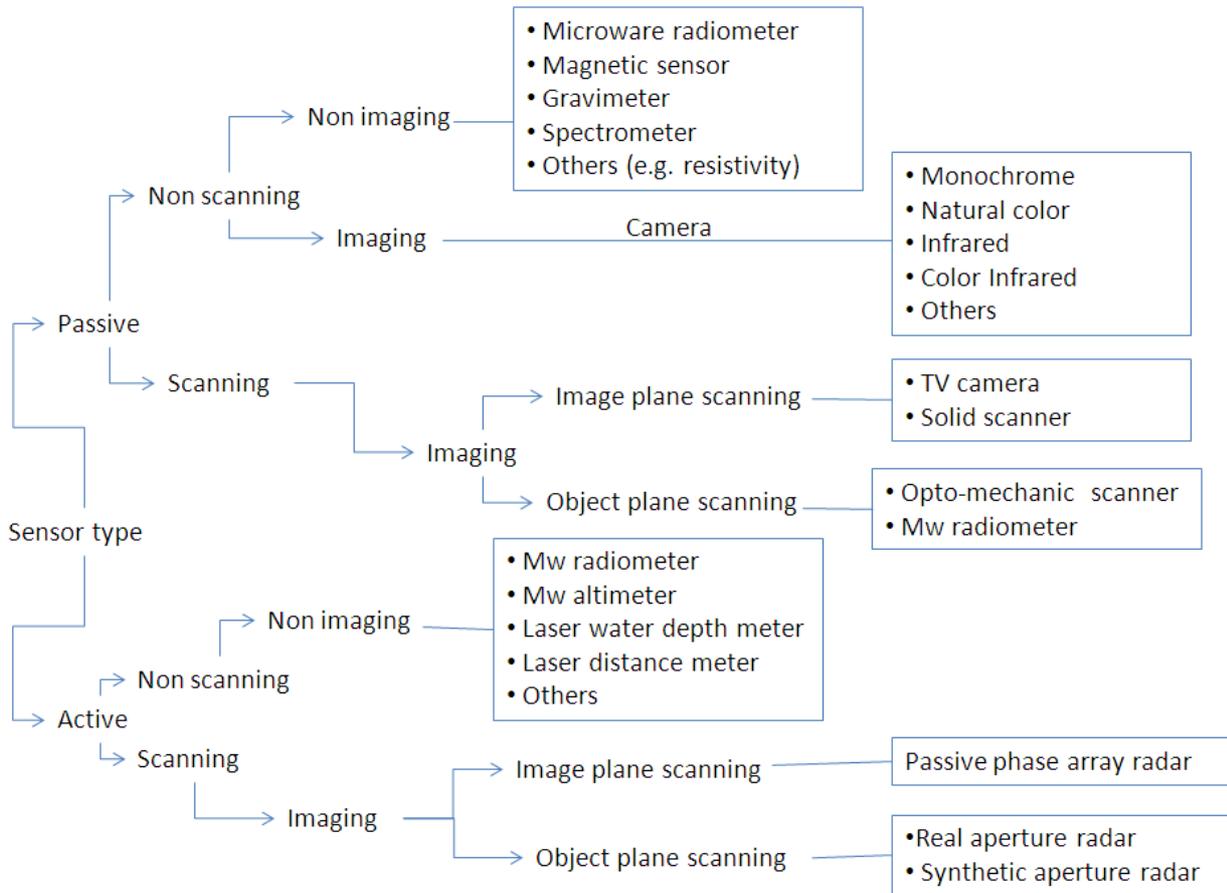


Figure 1 - EO related sensor taxonomy



Description cat.	CAT.	Sub-cat.	Description Sub-cat.	Spatial Resolution (Range)	Bands
SAR (High/Very high resolution)	C1	C1_1	Very high resolution	<1m + 3m	X
		C1_2	High resolution	3m + 25m	X,C,L
		C1_3	Medium resolution	25m + 50m	C,L
		C1_4	Low resolution	50m + >70m	C,L
Optical (High/Very High Resolution Multispectral)	C2	C2_1	Very high resolution multispectral	<1m + 8m	VIS,NIR
		C2_2	High resolution multispectral	8m + 30m	VIS,NIR
		C2_3	Hyperspectral	20m + 30m	VIS,NIR
		C2_4	Multispectral	30m + 90m	VIS,NIR,TIR
		C2_5	Panchromatic	<1m + 15m	Pan
		C2_6	SWIR	>20m + 70m	SWIR
		C2_7	Vis/IR	50m + 2700m	VIS,NIR,TIR
		C2_8	MW (passive)	23km + 32km	K,Ka
Medium resolution Land/ Ocean	C3	C3_1	Radiometers MultiSpectr.Imaging (MW)	1km + 100km	K,Ka,Ku
		C3_2	Radiometers MultiSpectr.Imaging (Vis/IR)	100m + 40km	VIS,NIR,SWIR,TIR,MWIR
		C3_3	Radiometers MultiSpectr.Sounding	48km	K->E
		C3_4	MultiDirection/MultiPolarization	6km + 47km	VIS,NIR
	C4	C4_5	Scatterometer	12km + 100km	C,Ku,L
		C4_6	Ocean colour instrument	236m + 825m	VIS,NIR
		C4_7	Altimetry	0.45m + 10km	Ku,Ka
		C4_8	Wind scatterometers	50km	C
		C4_9	Imaging multispectral radiometers (Vis/IR)	V: 5km H: 40+320km	UV,VIS,NIR,SWIR,TIR
		C4_10	Atmospheric chemistry	V: 1+132km H: 32+215km	UV,VIS,NIR,SWIR
		C4_11	Atmospheric temperature & humidity	V: 150m+3km H: 3+300km	TIR,SWIR, EHF
Atmospheric	C5	C5_1	Multiple direction/polarization radiometers	5.5km	VIS,NIR
		C5_2	Imaging multispectral radiometers (MW)	20+40km	K,Ka,W
		C5_3	Earth radiation budget radiometers	10km, 40km	UV,VIS,SWIR,FIR,TIR
		C5_4	LIDAR	V: 1+2km H: 300m	UV
		C5_5	Cloud profile & rain radar	H: 500m	W
		C5_6	Radio Occultation Sounder for the Atmosphere	res<1K	L
		C5_7	3-X gravity gradiometers		
		C5_8	GPS precision positioning		
		C5_9	Laser reflector (precise orbit)		
		C5_10	Sat-to-Sat tracking		
		C5_11	MIRAS		
Other scientific	C6	C6_1	Accelerometers		
		C6_2	Absolute scalar magnetometer		
		C6_3	Electric field		
		C6_4	Vector field magnetometer		
		C6_5	Temperature and water vapour		
		C6_6	Flux gate magnetometer		

Table 7 - Earth Observation sensed data classification

1.7 List of Earth Science fields and applications domains considered

The Earth science fields and applications domains considered in this document are grouped in the following list:

Atmosphere

- Long term NO₂ /O₃
- UV Radiation
- Greenhouse gases
- Aerosol
- Water vapour
- Stratospheric constituents
- Pollution
- Volcanic emissions
- Flights' vapour trails
- Clouds
- Air quality

Biosphere

- Carbon cycle and biomasses
- Coastal erosion, subduction
- Coastal vulnerability
- Fire monitoring & atlas
- Land use and land coverage
- Volatile organic compounds
- Forestry
- Global vegetation
- Bio-geophysical products

Glaciology (Cryosphere)

- Glaciers and ice-shelf
- Icebergs & ice moving platforms
- Long term study of the ice cycle
- Ice shelf extension and thickness
- Polar Regions
- Permafrost
- Ice-sea
- Cryosphere & climate
- Snow

Oceanography & Hydrology

- Biological oceanography
- Chemical oceanography
- Physical oceanography

Geophysics

- Geology & Geomorphology

- Geophysics
- Geodesy
- Magnetic field
- Gravity field
- Earthquake & Tectonics
- Volcanoes
- Soil science
- Landslides

Earth as system

- Earth System Science
- Water variability
- Desertification
- Transcontinental air pollution
- Shifts in ecosystem structure
- Human health and climate change
- Hazards assessment & mapping

Calibration & Validation (of satellite based sensors/instruments/missions)

- OPT/IR
- Altimetry
- Atmosphere
- SAR
- MW Radiometer

Geoinformatics

- Spatial based decision support systems
- Emergency support systems
- Archaeology
- Resource management
- Human geography
- Cartography
- Geo-statistics
- DEM

Requirements have been collected from users pertaining to the different fields and application domains listed above as highlighted in Figure 2.

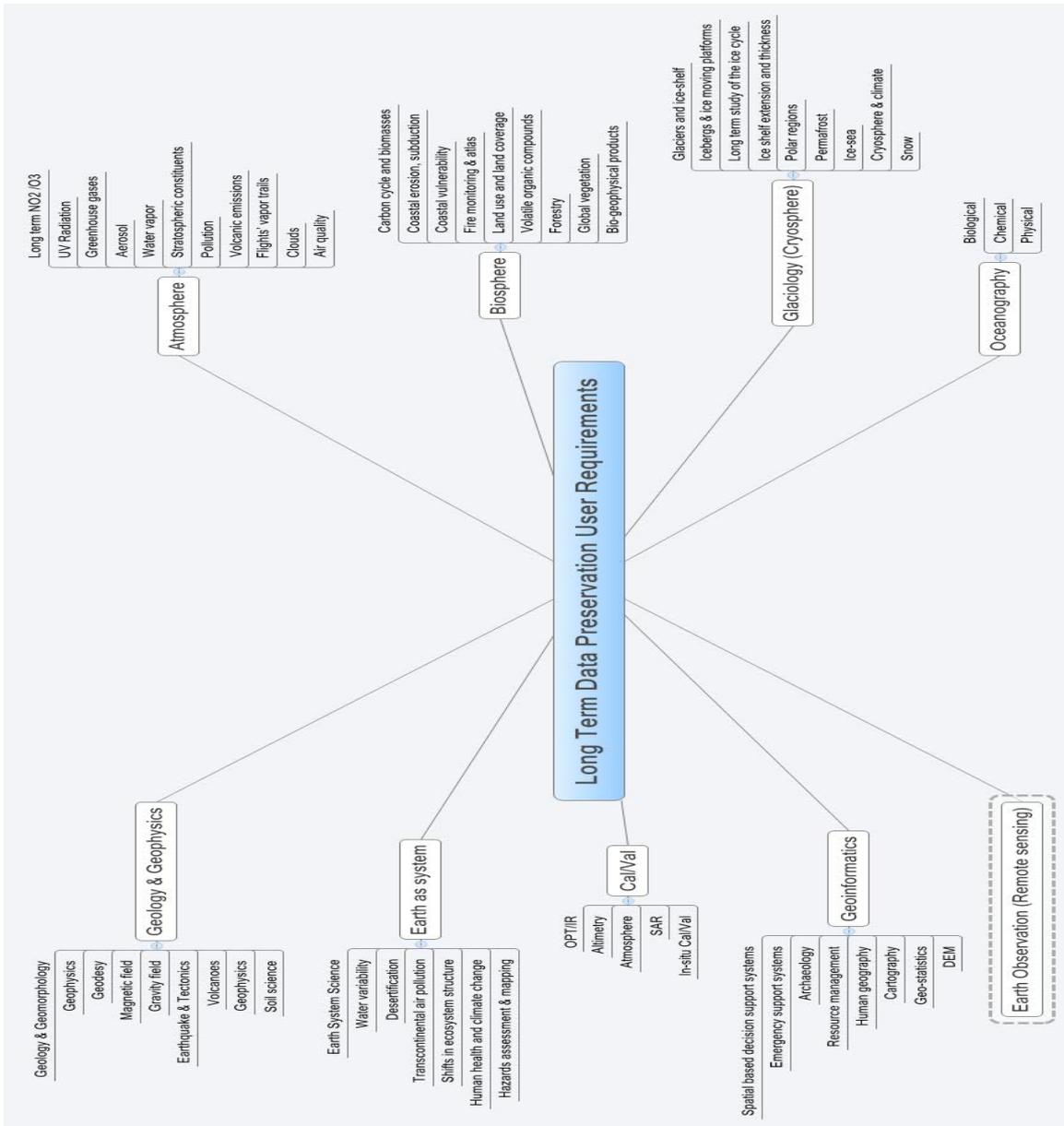


Figure 2 - FURD Earth Science domains for requirements context



1.8 List of international initiatives and programmes analysed

The list of international initiatives considered and analysed in the study to gather requirements and needs in terms of long term data preservation is provided in Table 3 and Table 4.

1.9 Campaigns

Table 8 contains the Earth Science campaigns which information have been retrieved and used to compare and understand some of the requirements listed in this document. Have been used documents and synthesis freely available via internet and/or magazines and bulletins.

Campaign	Purpose and other information
ESA and/or ESA participated campaigns	ESA EO Portal The datasets resulting from ESA airborne campaigns, available on Internet or media URL http://earth.esa.int/campaigns/index.htm
<i>DLR campaigns</i>	ADM-Aeolus prelaunch http://www.dlr.de/en/desktopdefault.aspx/tabid-1/86_read-23312/ Tandem-X http://www.dlr.de/eo/en/desktopdefault.aspx/tabid-5719/10058_read-24034/ , http://europlanet.dlr.de/node/index.php?id=492
CNES campaigns (balloons only)	Balloon campaigns: http://www.cnes.fr/web/CNES-en/8432-gp-new-stratospheric-balloon-campaign-starts-in-kiruna-on-1-april.php (search in local web site)
AEROCONTRAIL	The AEROCONTRAIL project aimed at a better understanding of the effect of aircraft exhaust on contrail and aerosol formation and of their impact on the climate. This project which included the combination of ground-based observations by Lidar, aircraft in situ observations (DLR-Falcon aircraft) and numerical modelling involved different participants: Stockholm University, DLR (FRG), IFU (FRG), ONERA, University of Strasbourg/CNRS (LPCA), University Blaise Pascal/CNRS (LaMP).
AGRISAR 86	SAR data collection during growing season (European crop coverage). ESA & JRC. Dated 1986, Airborne SAR. Data availability unknown.
AGRISAR 2006	The AGRISAR 2006 campaign was collecting in-situ, airborne SAR and optical measurements in support of decisions being taken on satellite instrument configurations for the first Sentinel Missions, as well as providing an important database for the study of longer term mission concepts. DLR, 2006. URL:

	http://www.dlr.de/hr/desktopdefault.aspx/tabid-4382/7152_read-10773/
AGRISCATT	Radar data acquisition at agricultural sites in five West-European countries during the 1987 and 1988 growing seasons. ESA/JRC, 1998.
ALPEX	Part of the Global Atmospheric Research Program (up to 1983)
Airborne Polar Experiment (APE)	APE is an international scientific programme. The M-55 Geophysica stratospheric aircraft flies with scientific instruments onboard to study the chemistry-physics of the stratosphere. URL: http://ape.ifac.cnr.it/the_campaigns.htm
BACC , BALTEX, BASIS, BASYS, BED, BEERS, BEPERS, BMP, BOSEX	Baltic Sea related campaigns and experiments. Campaigns and experiments are concerned with climate change, sea-ice relation and ice cycle over Baltic area. See www.geophysica.fi
CASP	Canadian Atlantic Storms Program, 1988. The goals of CASP were to begin the process of understanding and eventually better predicting the mesoscale structure of East Coast storms as well as the storms themselves. Conceptual models of the storms have been formulated, the nature of cyclogenesis and the structure of frontal surfaces have been investigated, and precipitation regions and precipitation type transitions have been studied. Ref. http://journals.ametsoc.org/
DAISEX	Digital airborne imaging spectrometer experiment, ESA 1999. It is the purpose of this experiment to evidence the feasibility of quantitatively retrieving geo/biophysical variables by controlling atmospheric effects and analyzing at the same time possible additional information present in directional anisotropy. Ref. ESA Bulletin 105 February 2001.
EASOE SESAME	European Arctic Stratospheric Ozone Experiment 1991-1992 Second European Stratospheric Arctic and Mid-latitude Experiment (SESAME, 1994-95) Ref. www.wmo.int and MIPAS instrument preparatory phase
EFEDA	European Field Experiment in Desertification threatened Area, Within the framework of the European Field Experiment in a Desertification-threatened Area (EFEDA) theoretical studies and intensive field campaigns have been carried out in order to model and measure the surface energy balance (SEE) and water transfer processes between soil, vegetation and the atmosphere in semiarid areas, where water availability is a limiting condition. Ref. IEEE Xplore. Sat data from: Landsat-TM, NOAA-AVHRR, Meteosat, ERS-2. Additional data from different sources.
ESABC	The Envisat Stratospheric Aircraft and Balloon Campaign (ESABC) Ref. ESA, Earth Science Division ESTEC URL: http://www.esa.int/esaEO/ESABLMZPD4D_index_0.html http://www.esa.int/esaEO/ESASEF7708D_index_0.html
IMAGES	International Marine Global Change Study, ref. http://www.globec.org/ , http://www.igbp.net/ , Bremen University http://www.marum.de/en/
SPARC	ESA Spectra-Barrax Campaign (SPARC), Barrax (Spain) July 2003 See CHRIS/Proba Workshop, URL: http://www.uv.es/leo/sparc/
General and reports	GAO/USA, GAO Report on Data Sharing in Climate Science, September 2007 (URL)



	http://republicans.energycommerce.house.gov/Media/File/News/10.2.07_GAO_Report_Data_Sharing_Climate_Research.pdf
--	---

Table 8 - Campaigns

2. REQUIREMENTS

2.1 Requirements Classification

Requirements are presented in a separate section for each Earth Science field and application domain. They have been classified in the study in relation to the category of Earth Science data to which they are addressed. Each single requirement is uniquely identified by the code:

FURD-XX-nnnn

where XX indicates the category of Earth Science data to which the requirement is addressed according to the following list:

- ES = “Earth Science”: requirements addressed to all the different categories of data (from C1 to C9) and applicable to all the associated data providers.
- EO = “Earth Observation”: requirements addressed to the different categories of data composing the Earth Observation data (from C1 to C5) and applicable to all the associated data providers.
- AB = “Airborne and Balloon”: requirements addressed to the C6 and C7 categories of data and applicable to all the associated data providers.
- IS = “In Situ”: requirements addressed to the C8 and C9 categories of data and applicable to all the associated data providers.
- CO = Common requirements, pertaining to all the categories of data (from C1 to C9) but also common to all the different Earth Science domains. These requirements are the highest level ones.
- ZZ = additional requirements, being requirements without a specific classification and always to be considered i.e. in addition to already provided ones.

and *nnnn* is a number unique within this document.

Each requirement has also associated a reference to its source as follows:

- **Interviews**: Interviews or questionnaires; within this document, requirements having source from interview and/or questionnaires are all identified as Interviews. Direct traceability to the interviewee is not allowed for privacy rule.
- **SOR-xx**: identifies the document analysed and containing a statement associable as needs or requirements or clear requirements.
- **Study**, when the requirement capture process is capable to sort out the existence of a need from one or many different sources (SOR, Interviews, RTS).

The following sections where requirements are presented are organized according to the following schema:

- Description of the Earth Science domain.
- General data needs: defining what kind of data is needed in the specific application domain but still not containing requirements.



- Earth Science Data User Requirements: requirements applicable to all ES data categories (from C1 to C9).
- Earth Observation Data User Requirements: requirements applicable to all EO data categories (from C1 to C5).
- Additional Data User Requirements: additional requirements applicable to Airborne, Balloon and In-Situ data categories (from C6 to C9).
- Description and characteristics of Earth Science data needed with specific detail relevant to Dataset composition:
 - Earth Observation products category (C1,C2,C3,C4,C5):
 - Resolution (medium, high, very high).
 - Band (when necessary).
 - Other data/products and categories (C6, C7, C8, C9).
 - Type of product (primary, secondary).

As an example, the requirements applicable to the Earth Observation data are all the ones labelled with ES (applicable to ES data) plus the ones labelled with EO (applicable to EO data).

2.2 Survey Population Statistics

More than 1200 persons have been invited to participate to the study activities through the compilation of questionnaires (a short generic one addressing LTDP at high level and a second one more specific) or through interviews.

The generic questionnaire has been included in the ESA User Services Satisfaction Survey mainly performed at the Bergen Living Planet Symposium 2010 (Norway, Bergen, from 28 June to 2 July 2010) and has included generic questions on possible interest and need in terms of data preservation.

The specific questionnaire was build with inquiries aimed at identify the specific earth science, its data context including in-situ data and cross utilisation of different data, and concerned interests in preservation. This was organised in order to capture details about types and utilisation of scientific data/products and main expectations about their preservation.

The questions were grouped in five categories:

1. Demographics and types of interests (including all different scientific interests).
2. Perception of the importance of data preservation and its motivation.
3. Which kind of information should be preserved: granularity and level of abstraction (both for primary and secondary data including in-situ of all types C1..C9).
4. How long would be preserved and typical time-window/time-series.
5. Other questions concerned additional data to be preserved.

Most questions offered multiple-choice answers. In addition several allowed open answers.

More than 120 persons have replied to the specific questionnaire or have been interviewed, some 170 have replied to the generic questionnaire at the Bergen Symposium. Fifteen interviews have been performed with scientist and specialist of specific fields; some other people (25) have provided comments or points of attention by a call or email on their autonomous decision.

The overall population to be accounted for the capture of information and requirements is around 310 people.

The following Figure 3 shows distribution of population with respect to the different scientific domains and disciplines. It is easy to see that all different scientific disciplines are present and significantly represented.

Furthermore, it is to be noted that several questionnaires show the same researcher involved or having interests in different projects, programmes and scientific disciplines. This spread interest has impacts on utilisation of similar scientific data or products in different context. This situation is common in all scientific context analysed (ref. Table 3).

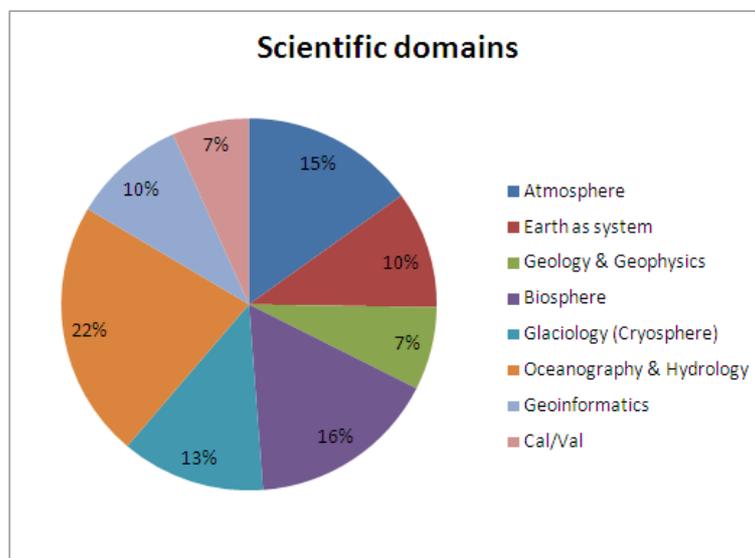
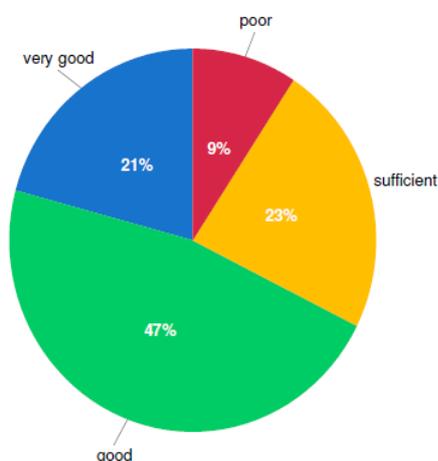


Figure 3 - Scientific population distribution in the LTDP/FIRST survey

Overall Satisfaction with Long Term Data Availability (n=173)



Importance of Present Availability of LTD (n=171)

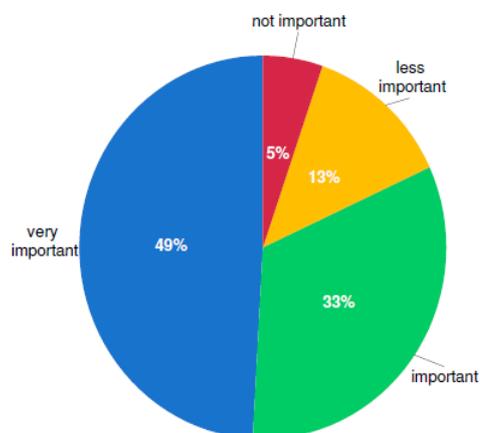


Figure 4 - Generic survey (from ESA User Service Survey at Bergen symposium)

2.3 Common requirements

Here are listed the requirements that have been identified as common to all Earth Science domains and applications. These requirements have been indicated by almost all interviewees in the survey, or it is possible to identify them in all documents considered as source (see Table 3), and do not have a specific characterisation to a specific Earth Science domain or application. Hence the following requirements are considered independent with respect to the any of the Earth Science domain or application. Those requirements are considered common and identified by FURD-CO-xxxx. Their applicability is on top of all other requirements.

ID	Description	Source
FURD-CO-0010	Access to data and information including documentation for scientific purposes should be free and open.	Interviews, SOR-1 (*)
FURD- CO -0020	The owners or providers of Earth Science data and information including documentation should guarantee their preservation without limits (all forever).	Interviews, SOR-1 (*)
FURD-CO-0030	Documents must be aligned (**) with and matching to models, algorithms, procedures and data versioning (coherence of information principle).	Interviews, SOR-1, 3, 4, 5
FURD-CO-0040	Access mechanism to data should be simple, easily available, easily deployable, and economical for the user.	Interviews, SOR-1, 3, 4, 5
FURD-CO-0050	Data and information integrity, quality and reliability should be guaranteed and documented by the owner or provider.	Interviews, SOR-1, 2, 3, 4, 5, 6
FURD-CO-0060	Data, products and information should be made available on request at any time.	Interviews, SOR-1, 2, 3, 4, 5, 6
FURD-CO-0070	LTDP must define homogeneous conditions (***) of preservation. Homogeneous conditions of preservation shall be guaranteed by data providers.	Interviews, SOR-1,3
FURD-CO-0080	Information concerned with reference, provenance (****), context, fixity and access rights or conditions should be provided to the user by the data provider.	Interviews, SOR-3, 4

Note of the writer:

(*) SOR-1 recalls the necessity to grant access to data and their full availability; does not address in detail the cost aspects.

(**) Alignment concept is concerned with coherent description of all aspects of a topic. It identifies the structural relations among elements like: physical/chemical model, scientific and technical documentation, developed/changed algorithms, data acquisition and concerned transformation procedures. Due to somewhat different nature of different element but also due to historical reasons and different lifecycle of the elements, the spot at a specific time could highlight differences at first look. This prone to use configuration-like policies to maintain the alignment at a specific time and conditions.

(***)The concept of “**homogeneous conditions**” is quite wide. Nevertheless in this document the following points are addressed:

1. Homogeneity in the dataset means internal coherence of components of the data set. Homogeneous data are drawn from a single population or acquisition process. In other words, all outside processes that could potentially affect the data must remain constant for the complete time period of the sample. Inhomogeneities are caused when changes affect the acquisition/transformation process of one or more elements of the dataset or properties of the observations through time. These changes may be abrupt or gradual, depending on the nature of the disturbance (e.g. instrument subtle failure, or calibration change, or algorithm’s implementation changes). Realistically, obtaining perfectly homogeneous data is almost impossible, as unavoidable changes in the process of observing station will often affect the data.
2. Homogeneity among datasets means usability of similar datasets provided by different sources. This is related to different aspects due to the applied technologies, systems or procedures to capture and maintain data.
3. Homogeneity on temporal scale identifies the continuity in using similar datasets in future.

Furthermore, homogeneity includes the concept of (dis)solution of continuity in a series of observation or in general of information. Homogeneity is recalled in SOR-1 GCOS-138 document and is a basic concept for ECVs and FCDRs (ref. Table 2: SOR-1, SOR-6).

(***) The concept of **provenance** is "to come from", means the origin or the source of data or information, the history of the ownership and location of the object. Furthermore, the concept of **provenance** is to be considered as follow:

- a) In case of human produced data or information, references to the person and/or organisation must be provided
- b) In case of machine produced data or information, reference to the algorithm, procedure, and/or instrument generating data must be provided.

2.4 Atmosphere Domain

2.4.1 Domain Description

Atmospheric sciences cover the gaseous part of the Earth (the atmosphere) between the surface and the exosphere (some 1000 Km above surface) . According to the general accepted division, considered sub disciplines are:

- Atmospheric physic and chemistry
- Climatology
- Meteorology
- Paleoclimatology

In the current study we have **not considered** Paleoclimatology, not having yet evidences of utilisation of long term data retention and/or Earth Observation remote sensed data or similar products used by this scientific community.

We have considered and analysed requirements directly issued or retrievable from documents and relevant to the application fields:

1. **Long term NO₂/O₃.** Earth's atmosphere is made up of a combination of gases. The major components of nitrogen, oxygen, and argon remain constant over time and space, while trace components like CO₂ and water vapour vary considerably over both space and time. This branch focuses on long term variation (or stability) of atmospheric gases.
2. **UV Radiation.** Earth's atmosphere blocks much of Sun emitted ultraviolet radiation, however the remaining portion causes a number of effects on atmosphere and people too. One of the most filtering atmospheric components is Ozone.
3. **Greenhouse gases.** These are the ones that in an atmosphere absorb and emit radiation within the thermal infrared range. This process is the fundamental cause of the greenhouse effect (thermal change) in the terrestrial atmosphere. The primary greenhouse gases in the Earth's atmosphere are water vapour, carbon dioxide, methane, nitrous oxide, and ozone.
4. **Aerosol.** Those are microscopic particle in the atmosphere. Sources of human-generated aerosol include industry, temperature conditioning, motor vehicles and vegetation burning. Natural sources include volcanoes, dust storms and ocean plankton. According to some recent studies, aerosol is responsible both for cooling and heating the planet.
5. **Water vapour.** Atmospheric Water Vapour, Precipitable Water Vapour, or Humidity is a measure of the water vapour content of the air. It is one of the most important greenhouse gas of the atmosphere. Water vapour is a major vehicle for the transport of energy (latent heat) and acts as a regulator of planetary temperatures through absorption and emission of radiation, most significantly in the thermal infrared (the greenhouse effect).

6. **Stratospheric constituents.** The stratosphere is a region of intense interactions among radiative, dynamical, and chemical processes, in which horizontal mixing of gaseous components proceeds much more rapidly than vertical mixing. TOA refers to Top Of Atmosphere (TOA) modelling and analysis too.
7. **Pollution;** here is intended as air/atmosphere pollution which is a chemical, particulate or biological agent that modifies the natural (common) characteristics of the atmosphere.
8. **Clouds;** Cloud shapes, and the dynamics of their formation, are accurate indicators of important atmospheric properties, including air stability, moisture content, and motion.
9. **Flights' vapour trails,** or contrails, affect the Earth's radiation balance acting as a radiative forcing. Studies have found that vapour trails or contrails trap outgoing longwave radiation emitted by the Earth and atmosphere (positive radiative forcing) at a greater rate than they reflect incoming solar radiation (negative radiative forcing).
10. **Air quality** field has the objective to analyse and predict the air quality for health and safety purpose.
11. **Volcanic emissions** field is concerned with studies and analysis of interaction between the atmosphere and gases and particles emitted by the volcano.

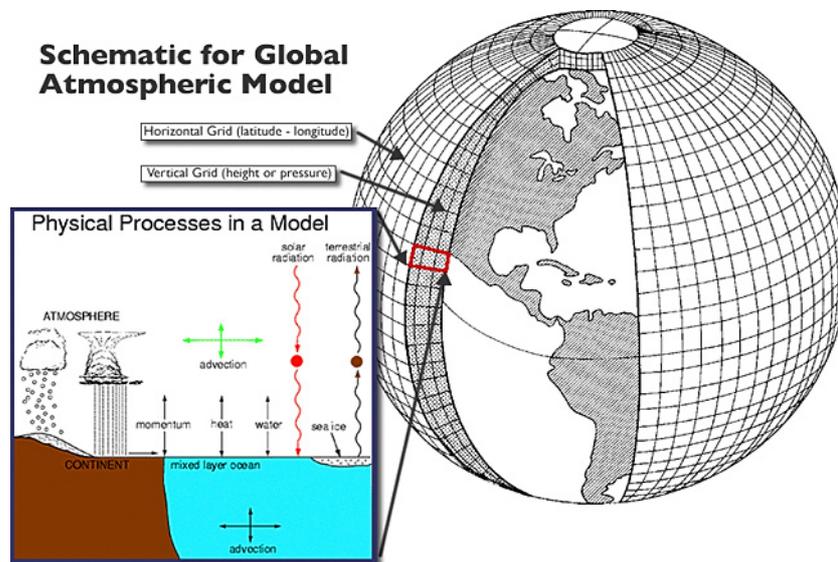


Figure 5 - Schematic for Global Atmosphere Circulation Model (from rst.gsfc.nasa.gov)

2.4.2 General data needs

Scientific data and information needs are concerned with:



1. Atmosphere models and references (e.g. General Circulation Model, Atmosphere Circulation Model ACM, etc.).
2. Atmospheric composition data (e.g. gases, particles, etc.).
3. Atmosphere physics and thermodynamics.
4. Long term time series for validation of models, analysis of trends and variations.

Needed types of datasets are relevant to:

1. Earth Observation space sensed data pertaining atmospheric physic and chemistry of atmosphere (C1, C2, C3, C4 and C5) both Primary and Secondary.
2. In-situ data (C6,C7) pertaining physic and chemistry of atmosphere at different layers.
3. In-situ data (C8,C9) pertaining physic and chemistry of atmosphere at low layers of interface with land and water.

2.4.3 Earth Science Data User Requirements

ID	Description	Source
FURD-ES-0010	It is required to have access to all data, information and documents, including models and algorithms.	Interviews, SOR-1,2,3,6, 7, 8, Study,
FURD-ES-0020	It is required the preservation of scientific data and relevant documentation for more than 20 (twenty) years.	Interviews, Study,
FURD-ES-0030	Complete cancellation of data and documents must be avoided.	Interviews
FURD-ES-0040	Time series/time windows of data should be greater than 20 years.	Interviews, Study,
FURD-ES-0050	Documents are required describing formats and representations of data.	Interviews, SOR-1
FURD-ES-0060	Analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-1,2,3, Study,
FURD-ES-0070	Analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5
FURD-ES-0080	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-0090	Various kinds of auxiliary data are required for atmospheric and geometric correction.	Interviews, SOR-3, 6
FURD-ES-0100	Mechanism to access data and information (documents) must be simple and easy.	Interviews, SOR-1,2,3, 4
FURD-ES-0110	Access to data and information should be free of charge	Interviews
FURD-ES-0120	Performance of data access mechanisms should be at technological frontier (*).	Interviews, SOR-1
FURD-ES-0130	Data owner must guarantee integrity and quality of data, products and documents without limits.	Interviews, SOR-2, 3

FURD-ES-0140	Data access security policies must not prevent or limit accessibility.	Interviews
FURD-ES-0150	Reprocessing shall be possible on request at any time and for any timespan.	Interviews, SOR-3
FURD-ES-0160	Temporal coherence of time series of data must be guaranteed or gaps must be listed clearly	Interviews, Study
FURD-ES-0170	Data and documents concerned with campaigns must be available for more than 20 years.	Interviews, Study,
FURD-ES-0180	Atmosphere analysis requires reprocessing of data.	Interviews, SOR-3
FURD-ES-0190	Analysis requires preservation of reprocessed data and products for 5 (five) years (at least).	Interviews
FURD-ES-0200	Atmosphere analysis requires continuous time series of data.	Interviews, SOR-3
FURD-ES-0210	Generated data and products must be traceable.	Interviews, SOR-3
FURD-ES-0220	Owner of C6, C7, C8, C9 data and products should grant their availability and exploitability at any time.	Interviews, SOR-3, Study

Notes of writer:

(*) This should be considered a wish even if required by many users and generically recalled in some documents (e.g. SOR-1) being not considered any associated cost model.

2.4.4 EO Data User Requirements

ID	Description	Source
FURD-EO-0010	Atmosphere analysis requires time series of data for more than 20 (twenty) years.	Interviews, SOR-2, SOR-3
FURD-EO-0020	EO data and products of all categories must be available as soon as possible.	Interviews, SOR-2, SOR-3
FURD-EO-0030	Atmosphere analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types.	Interviews, SOR-2, SOR-3
FURD-EO-0040	Atmosphere analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5.	Interviews, SOR-2, SOR-3
FURD-EO-0050	Atmosphere analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, C3-3.	Interviews, SOR-2, SOR-3
FURD-EO-0060	Atmosphere analysis requires Primary Data of type C4 of all sub-categories.	Interviews, SOR-2, SOR-3
FURD-EO-0070	Atmosphere analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10.	Interviews, SOR-2, SOR-3
FURD-EO-0080	Atmosphere analysis requires SAR (VHR/HR/MR) from tandem/combined missions.	Interviews, SOR-2, SOR-3
FURD-EO-0090	Atmosphere analysis requires OPT/Multispectral from combined missions including stereo.	Interviews, Study, SOR-2
FURD-EO-0100	Atmosphere analysis requires raw and higher product (i.e. L0,L1,L2, L3).	Interviews
FURD-EO-0110	Atmosphere analysis requires secondary data: Ancillary, Auxiliary, CAL/VAL.	Interviews, SOR-1,2, 3
FURD-EO-0120	Atmosphere analysis requires metadata and browse images for data discovery.	Interviews, SOR-1,2, 3

FURD-EO-0130	Atmosphere analysis requires Mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3
FURD-EO-0140	Atmosphere analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-0150	Atmosphere analysis requires EO remote sensed data coherent with campaigns, at least temporarily.	Interviews, SOR-1, 2, 3
FURD-EO-0160	Atmosphere analysis requires access to all available EO data free of charge.	Interviews, SOR-1, 2, 3

2.4.5 Additional Data User requirements

ID	Description	Source
FURD-AB-0010	Atmosphere analysis requires campaign data and products (i.e. C6, C7) for comparison, correction and integration to satellite remote sensed data (i.e. EO).	Interviews, Study
FURD-AB-0020	Atmosphere analysis requires campaign related (i.e. C6, C7) documents aligned with C6/C7 data and product versions.	Interviews, Study
FURD-AB-0030	Atmosphere analysis requires campaign related (i.e. C6, C7) data and products which timeframe is coherent with EO remote sensed data.	Interviews, Study
FURD-AB-0040	Owner of C6, C7 data and products should grant data availability and exploitability at any time.	Interviews, Study
FURD-AB-0050	Atmosphere analysis requires in-situ data and products (i.e. C8, C9) for comparison, correction and integration to satellite remote sensed data (i.e. EO).	Interviews, Study
FURD-AB-0060	Atmosphere analysis requires in-situ related (i.e. C8, C9) documents aligned with C6/C7 data and product versions.	Interviews, Study
FURD-AB-0070	Atmosphere analysis requires in-situ related (i.e. C8, C9) data and products which timeframe is coherent with EO remote sensed data.	Interviews, Study

2.4.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C1 (SAR, both sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C3 (Radiometric, all sub-types)
 - i. Raw data
 - ii. Higher level
 - d. C4 (Atmosphere Chemistry, all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)

- i. Higher level
 - f. C6 (Airborne)
 - g. C7 (Balloon)
 - h. C8 (Ground)
 - i. C9 (Hydro)
- 2. Secondary data:
 - a. Ancillary data
 - b. Auxiliary data to process telemetries
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
- 3. Metadata
- 4. Browse images

2.4.7 Special cases

- A. Cloud application represents a specific case:
 - 1. Being necessary to preserve raw (i.e. source packets) and L0 data/products being possible the execution of different/specific algorithms. So that “original data” must be available at any time
 - 2. Even if it is generically expected a long-term preservation, another common period of analysis (time series) is around 5 (five) years. Up to now are not known projects comparing new series towards older periods.
 - 3. Cloud (together with Water vapour) represents the most demanding “consumer” of secondary data in the Atmosphere domain
- B. Flights' vapour trails represents a special condition due to the fact that:
 - 1. From available evidence it seems interested to OPT/Multispectral data/products of raw and upper levels.
 - 2. The application looks at short term preservation (1-5 years) but is asking for long term data too
- C. Optical data/products presenting cloud coverage greater than the ratio expected by the user could be of interests for researchers working on cloud analysis and studies. This means that data/products having bad quality indicators due to excess of cloud coverage should not considered as trash but eligible for other uses.

2.5 Biosphere

2.5.1 Domain Description

The biosphere is the ensemble of ecosystems of the Earth. It is also identified as the zone of life on Earth. From the broadest bio-physiological point of view, the biosphere is the entire ecological system integrating all living beings and their relationships, including their interaction with the elements of the lithosphere, hydrosphere and atmosphere.

According to the Encyclopaedia Britannica it is the relatively thin life-supporting stratum of the earth's surface, extending from a few miles into the atmosphere to the deep-sea vents of the oceans. The biosphere is a global ecosystem that can be broken down into regional or local ecosystems, or biomes. Organisms in the biosphere are classified into trophic levels (see food chain) and communities.

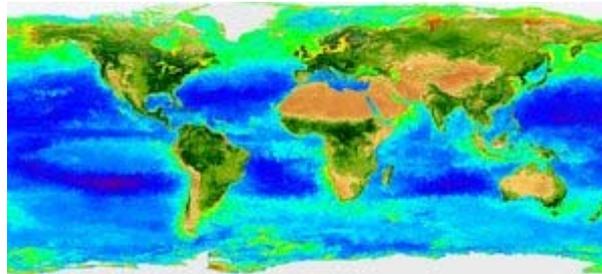


Figure 6 View of biosphere from remote sensing (from eoearth.org)

We have considered and analysed requirements directly issued or retrievable from documents and relevant to the application fields:

1. **Carbon cycle and biomasses**, the biogeochemical cycle by which carbon is exchanged among the biosphere, pedosphere, geosphere, hydrosphere, and atmosphere of the Earth. It is one of the most important cycles of the earth and allows for carbon to be recycled and reused throughout the biosphere and all of its organisms, thus including biomasses. The scientific field studies the cycle and its implications on resources as well as on climate change.
2. **Coastal erosion and subduction**. This scientific field focuses and analyses changes concerned with coastal alteration due to different phenomena and causes and its impacts on biosphere in general and critical systems like wetlands. It analyses complex combination of natural processes and manmade alterations.
3. **Coastal vulnerability**. Coastal zones are amongst the most dynamic natural environments on earth, providing a range of goods and services that are essential to human social and economic well-being. Coastal zones represent the narrow transitional zone between the world's land and oceans, characterized by highly diverse ecosystems such as coral reefs, mangroves, beaches, dunes and wetlands. Study of coastal vulnerability is becoming even more a must for a sustainable development.

4. **Fire monitoring and atlas.** In many vegetation types of the globe fire applied in agriculture and pastoralism and the occurrence of natural wildfires (natural fire regimes) are established elements in natural ecosystem processes and sustainable and productive traditional land-use systems. Excessive application of fire due to rapid demographic and land-use changes, however, increasingly lead to destruction of productivity, carrying capacity, biodiversity and vegetation cover. Climate variability such as periodic extreme droughts caused by different phenomena adds to the severity of fire impacts. Projected demographic and climate change scenarios suggest that this situation will become more critical during the next decades. The state of fire science (fundamental fire research, fire ecology) in vegetation types, and the results of biogeochemical and atmospheric sciences research provide knowledge for supporting decision making at many levels.
5. **Land use and coverage.** The Earth's biosphere produces many useful biological products for humans, including (but far from limited to) food, wood, pharmaceuticals, oxygen, and the recycling of many organic wastes. The land-based ecosystem depends upon topsoil and fresh water, and the oceanic ecosystem depends upon dissolved nutrients washed down from the land. Humans also live on the land by using building materials to construct shelters. This scientific field helps understanding of the complex interaction and better utilization.
6. **Volatile organic compounds.** A volatile organic compound (VOC) refers to organic chemical compounds which have significant vapor pressures and which can affect the environment and human health. VOCs are numerous, varied, and ubiquitous. Although VOCs include both man-made and naturally occurring chemical compounds. VOCs are typically not acutely toxic but have chronic effects. Because the concentrations are usually low and the symptoms slow to develop, analysis of VOCs, their extension and their effects is a demanding area.
7. **Forestry.** Forests are complex ecosystems, comprised of thousands of organisms, both plants and animals, but also insects, fungi, algae, and bacteria. The way we manage our natural resources is based upon our understanding of how these organisms live and interact with each other. It is also based on our understanding of the non-living components of the forest environment, such as geology, soil, water (hydrology), fire, and climatology. The vast number of different factors that affect the growth, health, management and utilization of forest ecosystems make the scientific field complex.
8. **Global vegetation.** Vegetative land cover is an important variable in many Earth system processes. General circulation and carbon exchange models require vegetative cover data. Vegetation also represents an important natural resource for humans and other species. Quantifying the types and extent of vegetation is important for resource management and for challenging issues regarding land cover change and global climate change. One of the key vital signs of earth's vegetation is the total green leaf area for a given ground area. Knowing the total leaf area in a plant helps scientists determine how much water will be stored and released by an ecosystem, how much leaf litter it will generate, and how much photosynthesis is going on. It also helps scientists understand the flow of energy among the various layers of vegetation, the atmosphere, and the ground, which in turn affects climate.

9. **Bio-geophysical products.** The biosphere affects Earth System behaviour through biogeophysical and biogeochemical processes and through biological diversity. Biological processes are important for the global-scale biogeochemical cycling of carbon, nitrogen, phosphorus, silicon and iron.

2.5.2 General Data Needs

Scientific data and information needs are concerned with:

1. Biosphere models and references (e.g. Biosphere generic model BIOGEM, CASA, BIOMOSA, CSIRO/CABLE, GLOBIO, energy balance models, etc.).
2. Chemical composition data (e.g. gases, particles, etc.).
3. Physical processes (e.g. photosynthesis, heat exchange,...).
4. Long term time series for validation of models, analysis of trends and variations.

Needed types of datasets are relevant to:

1. Earth Observation space data pertaining atmospheric physic and chemistry of atmosphere (C1, C2, C3, C4 and C5) both Primary and Secondary.
2. In-situ data (C6,C7) pertaining physic and chemistry of atmosphere at different layers.
3. In-situ data (C8,C9) pertaining physic and chemistry of atmosphere at low layers of interface with land and water.

2.5.3 Earth Science Data User Requirements

ID	Description	Source
FURD-ES-0300	It is required to have access to all data, information and documents, including models and algorithms.	Interviews, SOR-1,2,3,6, 7, 8
FURD-ES-0310	It is required preservation of scientific data and relevant documentation for more than 20 (twenty) years.	Interviews
FURD-ES-0320	Complete cancellation of data and documents must be avoided.	Interviews
FURD-ES-0330	Time series/time windows should be greater than 20 years.	Interviews, Study,
FURD-ES-0340	Documents describing formats and representations of data are required.	Interviews, SOR-1
FURD-ES-0350	Biosphere analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research	Interviews, SOR-1,2,3

	or application.	
FURD-ES-0360	Biosphere analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5
FURD-ES-0370	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-0380	Various kinds of auxiliary data are required for atmospheric and geometric correction.	Interviews, SOR-3, 6
FURD-ES-0390	Mechanism to access to data and information (documents) must be simple and easy	Interviews, SOR-1,2,3, 4
FURD-ES-0400	Mechanism to access data and information should be free of charge for users	Interviews
FURD-ES-0410	Performance to access data should be at technological frontier	Interviews, SOR-1
FURD-ES-0420	Data owner must grant integrity and quality of data, products and documents without limits.	Interviews, SOR-2, 3
FURD-ES-0430	Data access security policies must not prevent or limit accessibility.	Interviews
FURD-ES-0440	Reprocessing shall be possible on request at any time and for any timespan.	Interviews, SOR-3
FURD-ES-0450	Temporal coherence of time series must be granted or gaps must be listed clearly.	Interviews, Study
FURD-ES-0460	Data and documents concerned with campaigns must be available for more than 20 years.	Interviews
FURD-ES-0470	Biosphere analysis requires reprocessing.	Interviews, SOR-3
FURD-ES-0480	Biosphere analysis requires preservation of reprocessed data and products for 5 (five) years (at least).	Interviews
FURD-ES-0490	Biosphere analysis requires continuous time series.	Interviews, SOR-3
FURD-ES-0500	Generated data and products must be traceable.	Interviews, SOR-3
FURD-ES-0510	Owner of C6, C7, C8, C9 data and products should grant their availability and exploitability at any time.	Interviews, SOR-3, Study

2.5.4 EO Data User Requirements

ID	Description	Source
FURD-EO-0200	Biosphere analysis requires time series of data for more than 20 (twenty) years	Interviews, SOR-2, SOR-3
FURD-EO-0210	EO data and products of all categories must be available as soon as possible	Interviews, SOR-2, SOR-3
FURD-EO-0220	Biosphere analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types	Interviews, SOR-2, SOR-3
FURD-EO-0230	Biosphere analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5	Interviews, SOR-2, SOR-3
FURD-EO-0240	Biosphere analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, C3-3	Interviews, SOR-2, SOR-3
FURD-EO-0250	Biosphere analysis requires Primary Data of type C4 of all sub-categories	Interviews, SOR-2, SOR-3
FURD-EO-0260	Biosphere analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10	Interviews, SOR-2, SOR-3
FURD-EO-0270	Biosphere analysis requires SAR (VHR/HR/MR) from tandem/combined missions	Interviews, SOR-2, SOR-3
FURD-EO-0280	Biosphere analysis requires OPT/Multispectral from combined missions including stereo	Interviews, Study, SOR-2
FURD-EO-0290	Biosphere analysis requires raw and higher product (i.e. L0,L1,L2, L3)	Interviews, Study
FURD-EO-0300	Analysis requires secondary data: Ancillary, Auxiliary, CAL/VAL	Interviews, SOR-1,2, 3
FURD-EO-0310	Biosphere analysis requires metadata and browse images	Interviews, SOR-1,2, 3
FURD-EO-0320	Biosphere analysis requires mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3
FURD-EO-0330	Biosphere analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-0340	Biosphere analysis requires EO remote sensed data coherent with campaigns, at least temporarily	Interviews, SOR-1,2, 3
FURD-EO-0350	Biosphere analysis requires access to all available EO data free of charge	Interviews, SOR-1,2, 3

2.5.5 Additional Data User requirements

ID	Description	Source
FURD-AB-0100	Biosphere analysis requires campaign data and products (i.e.	Interviews,

	C6, C7) for comparison, correction and integration to satellite remote sensed data (i.e. EO)	Study
FURD-AB-0110	Biosphere analysis requires campaign related (i.e. C6, C7) documents aligned with C6/C7 data and product versions.	Interviews, Study
FURD-AB-0120	Biosphere analysis requires campaign related (i.e. C6, C7) data and products which timeframe is coherent with EO remote sensed data	Interviews, Study
FURD-AB-0130	Owner of C6, C7 data and products should grant data availability and exploitability at any time	Interviews, Study
FURD-AB-0140	Biosphere analysis requires in-situ data and products (i.e. C8, C9) for comparison, correction and integration to satellite remote sensed data (i.e. EO)	Interviews, Study
FURD-AB-0150	Biosphere analysis requires in-situ related (i.e. C8, C9) documents aligned with C6/C7 data and product versions.	Interviews, Study
FURD-AB-0160	Biosphere analysis requires in-situ related (i.e. C8, C9) data and products which timeframe is coherent with EO remote sensed data	Interviews, Study

2.5.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C2 (Optical, all sub-types)
 - i. Higher level
 - ii. Raw data
 - b. C1 (SAR, both sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - c. C3 (Radiometric, all sub-types)
 - i. Raw data
 - ii. Higher level
 - d. C4 (Atmosphere Chemistry, all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)
 - i. Higher level
2. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries:
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
3. Metadata
4. Browse image

Note: Priority C2/C1 is from captured information (interviews, questionnaires) where the most common sensor considered is of type C2 of all bands and resolution (e.g. AVHRR, MODIS, etc), however C1 is widely considered too. One example is furnished by the number of answers relevant to SAR combined missions. Thus it has been changed the hierarchical priority of C1/C2.



2.5.7 *Special case*

None.

2.6 Glaciology (Cryosphere)

2.6.1 Domain Description

Glaciology is the study of glaciers, or more generally ice and natural phenomena that involve ice. Glaciology is an interdisciplinary earth science that integrates geophysics, geology, physical geography, geomorphology, climatology, meteorology, hydrology, biology, and ecology. The impact of glaciers on humans adds the fields of human geography and anthropology.

We have considered and analysed requirements directly issued or retrievable from documents and relevant to the application fields:

1. **Glaciers and ice-shelf.** This field focuses on monitoring and studying glaciers and ice-shelf at global level including measurement of their flows, thickness and evolution. It is linked to climate analysis, water cycle, sea-level changes and catastrophic flooding from glacial lakes. It covers both at seasonal level and long term trends;
2. **Icebergs and ice moving platforms.** It is concerned with ice detection and tracking and with changes due to their movement. It includes reduction of shipping hazards related to sea ice and icebergs motion that is hardly predictable. Furthermore climate-related changes in the cryosphere involve the impacts of a reduced ice cover on biological habitat and sea level rise on coastal native communities as well as biological impacts;
3. **Long term study of the ice cycle.** The causes of ice ages are not fully understood for both the large-scale ice age periods and the smaller ebb and flow of glacial–interglacial periods within an ice age. The consensus is that several factors are important: atmospheric composition (the concentrations of carbon dioxide, methane. There is evidence that greenhouse gas levels fell at the start of ice ages and rose during the retreat of the ice sheets, but it is difficult to establish cause and effect. Greenhouse gas levels may also have been affected by other factors which have been proposed as causes of ice ages, such as the movement of continents and volcanism;
4. **Polar Regions.** The branch of science has the objective to increase our understanding of the ice cover and its connection to the rest of the climate systems. History and projections of global climate suggest that the high-latitude ice-covered regions of the Earth, the Arctic and Antarctica, have high sensitivity to climate change. Among the reasons for this sensitivity is the positive albedo feedback associated with the warming/cooling and melting/formation of snow and ice. Other factors include the changes in thermohaline circulation and energy exchanges associated with the formation and melting of sea ice. Consequently, an understanding of the Earth's ice cover and its connections to the rest of the climate systems, is essential to understanding the past, present, and future behaviour of the Earth system as a whole;
5. **Permafrost.** The extent of permafrost can vary as the climate changes. Today, a considerable area of the Arctic is covered by permafrost (including discontinuous



permafrost). Overlying permafrost is a thin active layer that seasonally thaws during the summer. Plant life can be supported only within the active layer since growth can occur only in soil that is fully thawed for some part of the year. Permafrost part of the carbon cycle and can also be storage of carbon.

6. **Ice and sea.** Understanding and simulation of sea ice and sea dynamics suggest that some of the most marked responses to climate change would be expected to occur in the high latitudes, and in particular in the Arctic. Time series of sea-ice concentration data are critical for identifying interannual and decadal fluctuations that could point to the existence of significant changes in oceanic and atmospheric circulation
7. **Cryosphere and climate.** Principal goal of this scientific field is to assess and quantify the impacts that climate variability and change have on components of the cryosphere and vice versa and its overall stability, and the consequences of these impacts for the climate system as a whole.
8. **Snow.** Large areal extent of cryosphere is represented by snow. It is studied for its high spatial and temporal variability in properties, impacts both global/regional energy and water cycles, high reflectance, thermal insulation as well as storage of water.

2.6.2 General Data Needs

Scientific data and information needs are concerned with:

1. Atmosphere/ocean/cryosphere coupled models, regional models and references (models are still under discussions).
2. Discontinuities (small and large scales).
3. Albedo studies, models and measurements.
4. Physic and chemistry at local and global scale
5. Long term time series for validation of models, analysis of trends and variations

Needed types of datasets are relevant to:

1. Earth Observation space data pertaining atmospheric physic and chemistry of atmosphere (C1, C2, C3, C4 and C5) both Primary and Secondary.
2. In-situ data (C6, C7) pertaining physic and chemistry at different layers.
3. In-situ data (C8, C9) pertaining physic and chemistry at low layers of interface with land, water and underwater (including permafrost).

2.6.3 Earth Science Data User Requirements

ID	Description	Source
FURD-ES-0600	It is required to have access to all data, information and documents, including models and algorithms.	Interviews, SOR-1,2,3,6, 7, 8
FURD-ES-0610	It is required preservation of scientific data and relevant	Interviews

	documentation for more than 25 (twenty) years.	
FURD-ES-0620	Complete cancellation of data and documents must be avoided.	Interviews
FURD-ES-0630	Time series/time windows should be greater than 20 years	Interviews
FURD-ES-0640	Documents are required describing formats and representations of data.	Interviews, SOR-1
FURD-ES-0650	Glaciology analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-1,2,3
FURD-ES-0660	Glaciology analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5
FURD-ES-0670	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-0680	Various kinds of auxiliary data are required for atmospheric and geometric correction.	Interviews, SOR-3, 6
FURD-ES-0690	Mechanism to access data and information (documents) must be simple and easy.	Interviews, SOR-1,2,3, 4
FURD-ES-0700	Mechanism to access data and information should be free of charge.	Interviews
FURD-ES-0710	Performance to access data should be at technological frontier.	Interviews, SOR-1
FURD-ES-0720	Data owner must grant integrity and quality of data, products and documents without limits	Interviews, SOR-2, 3
FURD-ES-0730	Data access security policies must not prevent or limit accessibility.	Interviews
FURD-ES-0740	Reprocessing shall be possible on request at any time and for any timespan.	Interviews, SOR-3
FURD-ES-0750	Temporal coherence of time series must be granted or gaps must be listed clearly.	Interviews, Study
FURD-ES-0760	Data and documents concerned with campaigns must be available for more than 20 years.	Interviews
FURD-ES-0770	Glaciology analysis requires reprocessing.	Interviews, SOR-3
FURD-ES-0780	Glaciology analysis requires preservation of reprocessed data and products for 5 (five) years (at least).	Interviews
FURD-ES-0790	Glaciology analysis requires continuous time series.	Interviews, SOR-3
FURD-ES-0800	Generated data and products must be traceable.	Interviews, SOR-3
FURD-ES-0810	Owner of C6, C7, C8 and C9 data and products should grant their availability and exploitability at any time.	Interviews, SOR-3, Study

2.6.4 EO Data User Requirements

ID	Description	Source
----	-------------	--------

FURD-EO-0400	Glaciology analysis requires time series of data for more than 25 (twenty) years.	Interviews, SOR-2, SOR-3
FURD-EO-0410	EO data and products of all categories must be available as soon as possible.	Interviews, SOR-2, SOR-3
FURD-EO-0420	Glaciology analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types.	Interviews, SOR-2, SOR-3
FURD-EO-0430	Glaciology analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5.	Interviews, SOR-2, SOR-3
FURD-EO-0440	Glaciology analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, C3-3.	Interviews, SOR-2, SOR-3
FURD-EO-0450	Glaciology analysis requires Primary Data of type C4 of all sub-categories.	Interviews, SOR-2, SOR-3
FURD-EO-0460	Glaciology analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10.	Interviews, SOR-2, SOR-3
FURD-EO-0470	Glaciology analysis requires SAR (VHR/HR/MR) from tandem/combined missions.	Interviews, SOR-2, SOR-3
FURD-EO-0480	Glaciology analysis requires OPT/Multispectral from combined missions including stereo.	Interviews, Study, SOR-2
FURD-EO-0490	Glaciology analysis requires raw and higher product (i.e. L0,L1,L2, L3).	Interviews,
FURD-EO-0500	Glaciology analysis requires secondary data: Ancillary, Auxiliary, CAL/VAL .	Interviews, SOR-1,2, 3
FURD-EO-0510	Glaciology analysis requires metadata and browse	Interviews, SOR-1,2, 3
FURD-EO-0520	Glaciology analysis requires Mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3
FURD-EO-0530	Glaciology analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-0540	Glaciology analysis requires EO remote sensed data coherent with campaigns, at least temporarily.	Interviews, SOR-1,2, 3
FURD-EO-0550	Glaciology analysis requires access to all available EO data free of charge.	Interviews, SOR-1,2, 3

2.6.5 Additional Data User requirements

ID	Description	Source
FURD-AB-0200	Glaciology analysis requires campaign data and products (i.e. C6, C7, C8, and C9).	Interviews, Study
FURD-AB-0210	Glaciology analysis requires integration with satellite remote sensed data (i.e. EO).	Interviews, Study
FURD-AB-0220	Glaciology analysis requires EO remote sensed data related to the campaign timeframe (i.e. C6, C7, C8, and C9).	Interviews, Study
FURD-AB-0230	Owner of C6, C7, C8, C9 data and products should grant data	Interviews,

	availability and exploitability at any time.	Study
FURD-AB-0240	Glaciology analysis requires in-situ data and products (i.e. C6, C7, C8, C9) for comparison, correction and integration to satellite remote sensed data (i.e. EO).	Interviews, Study
FURD-AB-0250	Glaciology analysis requires in-situ related (i.e. C6, C7, C8, C9) documents aligned between them.	Interviews, Study

2.6.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C1 (SAR, all sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C3 (Radiometric, all sub-types)
 - i. Raw data
 - ii. Higher level
 - d. C4 (Atmosphere Chemistry, all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)
 - i. Higher level
 - f. **C6 (Airborne)**
 - g. **C7 (Balloon)**
 - h. **C8 (Ground)**
 - i. **C9 (Hydro)**
1. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
2. Metadata
3. Browse image

Note:

1) Earth Science domain Glaciology (or Cryosphere) has a long tradition concerned with In-situ data capture, analysis and exploitation thus types C6, C7, C8 and C9 are of very high importance.

2) A special attention is to be made to the altimetry data and relevant performances being pivot information necessary to the models both for monitoring and analysis. Another point of attention is concerned with legacy data (e.g. www.the-cryosphere-discuss.net/3/513/2009/, WMO global cryosphere watch, SAON, http://geo-cryosphere.org/meetings/uscanada08/presentations/Tues/key_igos_gcw.pdf).



2.6.7 *Special case*

Special cases are concerned with: Polar Regions, Ice-Sea and Cryosphere vs Climate where combined information are necessary. In particular, there is huge interest on 3D gravity measurements and precise GPS positioning.

2.7 Oceanography

2.7.1 Domain Description

Oceanography (or marine science) is the branch of Earth science that studies the ocean. It covers a wide range of topics, including marine organisms and ecosystem dynamics; ocean currents, waves, and geophysical fluid dynamics; plate tectonics and the geology of the sea floor; and fluxes of various chemical substances and physical properties within the ocean and across its boundaries. These diverse topics reflect multiple disciplines that oceanographers blend to further knowledge of the World Ocean and understanding of processes within it: biology, chemistry, geology, meteorology, and physics as well as geography. Not all are covered in this document.

This section generically addresses the three main branches:

1. **Biological oceanography**, as the scientific study of organisms in the ocean or other marine or brackish bodies of water.
2. **Chemical oceanography**, as the scientific of the chemistry of the ocean and its chemical interaction with the atmosphere.
3. **Physical oceanography**, or marine physics, studies the ocean's physical attributes including temperature-salinity structure, mixing, waves, internal waves, surface tides, internal tides, and currents.

This document addresses requirements concerned with data captured using sensors of various kinds and does not take care of all the huge amount of work performed by ocean scientists. For example, here are not considered data preservation requirements that could be generated by scientists caring of protecting the world's oceans and the species that inhabit them.

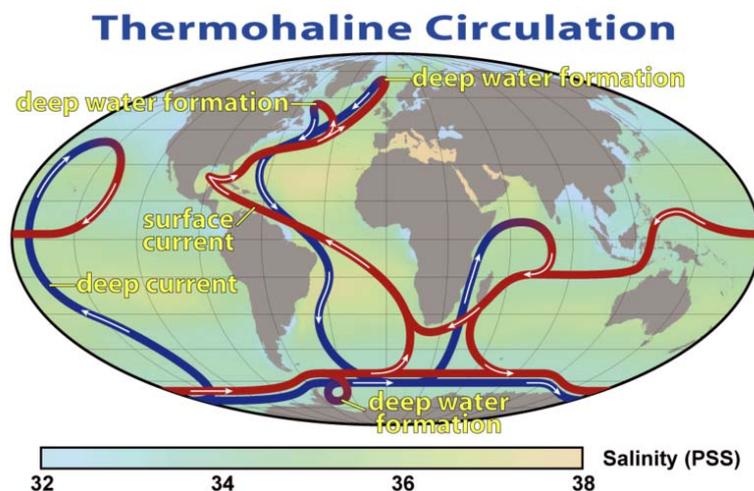


Figure 7 - View of thermohaline circulation (from maps grida no)

2.7.2 General Data Needs

Scientific data and information needs are concerned with:

1. Ocean models and references (e.g. General Circulation Model, Ocean Circulation Model(s), etc.).
2. Chemical composition and physical properties.

3. Bathymetric charts.
4. Marine biology.
5. Meteorology (forecast, history).
6. Meteorological models and meteorological data.
7. Long term time series for validation of models, analysis of trends and variations.

Needed types of datasets are relevant to:

1. Earth Observation space data pertaining atmospheric physic and chemistry of atmosphere (C1, C2, C3, C4 and C5) both Primary and Secondary.
2. In-situ data (C6, C7) pertaining physic and chemistry of atmosphere (close to water surface and just above) and ocean at different layers.
3. In-situ data (C8, C9) pertaining physic and chemistry of atmosphere at low layers of interface with land and water.

2.7.3 Earth Science Data User Requirements

ID	Description	Source
FURD-ES-0900	It is required to have access to all data, information and documents, including models and algorithms.	Interviews, SOR-1,2,3,6, 7, 8
FURD-ES-0910	It is required preservation of scientific data and relevant documentation for more than 20 (twenty) years.	Interviews
FURD-ES-0920	Complete cancellation of data and documents must be avoided.	Interviews
FURD-ES-0930	Time series/time windows should be greater than 20 years.	Interviews
FURD-ES-0940	Documents are required describing formats and representations of data.	Interviews, SOR-1
FURD-ES-0950	Oceanography analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-1,2,3
FURD-ES-0960	Oceanography analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5
FURD-ES-0970	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-0980	Various kinds of auxiliary data are required for atmospheric and geometric correction.	Interviews, SOR-3, 6
FURD-ES-0990	Mechanism to access data and information (documents) must be simple and easy.	Interviews, SOR-1,2,3, 4
FURD-ES-1000	Mechanism to access data and information should be free of charge.	Interviews
FURD-ES-1010	Performance to access data should be at technological frontier.	Interviews, SOR-1
FURD-ES-1020	Data owner must grant integrity and quality of data, products and documents without limits.	Interviews, SOR-2, 3
FURD-ES-1030	Data access security policies must not prevent or limit	Interviews

	accessibility.	
FURD-ES-1040	Reprocessing shall be possible on request at any time and for any timespan.	Interviews, SOR-3
FURD-ES-1050	Temporal coherence of time series must be granted or gaps must be listed clearly.	Interviews, Study
FURD-ES-1060	Data and documents concerned with campaigns must be available for more than 20 years.	Interviews
FURD-ES-1070	Oceanography analysis requires reprocessing.	Interviews, SOR-3
FURD-ES-1080	Analysis requires preservation of reprocessed data and products for 5 (five) years (at least).	Interviews
FURD-ES-1090	Oceanography analysis requires continuous time series.	Interviews, SOR-3
FURD-ES-1100	Generated data and products must be traceable.	Interviews, SOR-3
FURD-ES-1110	Owner of C6, C7, C8 and C9 data and products should grant their availability and exploitability at any time.	Interviews, SOR-3, Study

2.7.4 EO Data User Requirements

ID	Description	Source
FURD-EO-0600	Oceanography analysis requires time series of data for more than 15 (fifteen) years.	Interviews,
FURD-EO-0610	EO data and products of all categories must be available as soon as possible.	Interviews, SOR-2, SOR-3
FURD-EO-0620	Oceanography analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types.	Interviews, SOR-2, SOR-3
FURD-EO-0630	Oceanography analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5.	Interviews, SOR-2, SOR-3
FURD-EO-0640	Oceanography analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, C3-3.	Interviews, SOR-2, SOR-3
FURD-EO-0650	Oceanography analysis requires Primary Data of type C4 of all sub-categories.	Interviews, SOR-2, SOR-3
FURD-EO-0660	Oceanography analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10.	Interviews, SOR-2, SOR-3
FURD-EO-0670	Oceanography analysis requires SAR (VHR/HR/MR) from tandem/combined missions.	Interviews, SOR-2, SOR-3
FURD-EO-0680	Oceanography analysis requires OPT/Multispectral from combined missions including stereo.	Interviews, Study, SOR-2
FURD-EO-0690	Oceanography analysis requires raw and higher product (i.e. L0,L1,L2, L3).	Interviews,
FURD-EO-0700	Analysis requires secondary data: Ancillary, Auxiliary, CAL/VAL.	Interviews, SOR-1,2, 3
FURD-EO-0710	Oceanography analysis requires metadata and browse.	Interviews, SOR-1,2, 3
FURD-EO-0720	Oceanography analysis requires Mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3



FURD-EO-0730	Oceanography analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-0740	Oceanography analysis requires EO remote sensed data coherent with campaigns, at least temporarily..	Interviews, SOR-1,2, 3
FURD-EO-0750	Oceanography analysis requires access to all available EO data free of charge.	Interviews, SOR-1,2, 3

2.7.5 Additional Data User requirements

ID	Description	Source
FURD-AB-0300	Oceanography analysis requires campaign data and products (i.e. C6, C7, C8, C9)	Interviews, Study
FURD-AB-0310	Oceanography analysis analysis requires integration with satellite remote sensed data (i.e. EO)	Interviews, Study
FURD-AB-0320	Oceanography analysis analysis requires EO remote sensed data related to the campaign timeframe (i.e. C6, C7, C8, C9)	Interviews, Study
FURD-AB-0330	Owner of C6, C7, C8, C9 data and products should grant data availability and exploitability at any time	Interviews, Study
FURD-AB-0340	Oceanography analysis requires in-situ data and products (i.e. C6, C7, C8, C9) for comparison, correction and integration to satellite remote sensed data (i.e. EO)	Interviews, Study
FURD-AB-0350	Oceanography analysis requires in-situ related (i.e. C6, C7, C8, C9) documents aligned between them.	Interviews, Study

2.7.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data:
 - a. C1 (SAR, all sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C3 (Radiometric, all sub-types)
 - i. Raw data
 - ii. Higher level
 - d. C4 (Atmosphere Chemistry, all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)
 - i. Higher level
 - f. C6 (Airborne)
 - g. C7 (Balloon)
 - h. C8 (Ground)
 - i. C9 (Hydro)
2. Secondary data:
 - a. Ancillary data
 - b. Auxiliary data to process telemetries:
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
3. Metadata
4. Browse images

Note: Meteorological data and information are integral part of the oceanography science domain.

2.7.7 Special cases

None.

2.8 Geophysics

2.8.1 Domain Description

Geophysics is the study of the Earth by the quantitative observation of its physical properties. Geophysical data are used to observe tectonic plate motions, study the internal structure of the Earth, supplement data provided by geologic maps, etc. Geophysical survey data are used to analyze potential petroleum reservoirs and mineral deposits, to locate groundwater, to locate archaeological finds, to find the thicknesses of glaciers and soils, and for environmental management and remediation. The theories and techniques are employed extensively in the Earth sciences in general.

We have considered and analysed requirements directly issued or retrievable from documents and relevant to the application fields:

1. **Geology and geomorphology.** It is concerned with Earth Surface Processes (ESP) and Landforms and pertained phenomena and implications like the erosion, deposition and formation of landforms and sediments. Foci include the physical geography of rivers, valleys, glaciers, mountains, hills, slopes, coasts, deserts and estuaries environments.
2. **Geodesy.** Is the Earth science discipline dealing with the measurement and representation of the Earth, including its gravitational field, in a three-dimensional time-varying space. Here are studied geodynamical phenomena such as crustal motion, tides, and polar motion. For this purpose it is designed global and national control networks, using space and terrestrial techniques while relying on datums and coordinate systems.
3. **Geodynamics** (i.e. Earthquakes, landslides, tectonics, volcanoes). Geodynamics is a subfield of geophysics dealing with dynamics of the Earth. Geodynamics commonly use data from local sensors (e.g. gas, temperature), geodetic GPS, SAR, and seismology, along with numerical models, to study the evolution of the Earth's lithosphere, mantle and core.
4. **Soil science.** Soil science is the study of soil as a natural resource on the surface of the earth including soil formation, classification and mapping; physical, chemical, biological, and fertility properties of soils; and these properties in relation to the use and management of soils. Soil studies and researches are compelled by a need to understand soil in the context of climate change, greenhouse gases, and carbon cycle/retention.

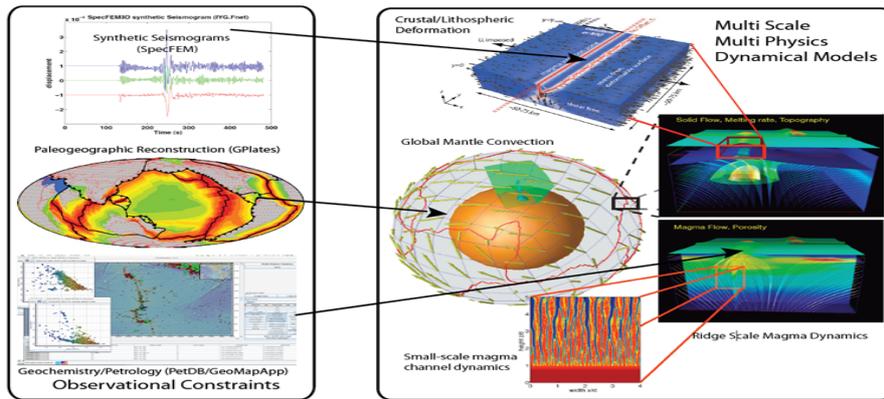


Figure 8 Geodynamic (from NSF)

2.8.2 General Data Needs

Scientific data and information needs are concerned with:

1. Physical phenomena like: gravity, heat flow, magnetism, radioactivity, fluid dynamics.
2. Tectonic and geodynamic models and references.
3. Volcanic related models and data.
4. Atmospheric composition data (e.g. electricity, ionosphere, chemistry, physics).
5. Water cycle models and data.
6. Long-term time series for validation of models, analysis of trends and variations.

Needed types of datasets are relevant to:

1. Earth Observation space data pertaining physic and chemistry (C1, C2, C3, C4 and C5) including magnetic and gravitational fields; both primary and secondary.
2. In-situ data (C6, C7) pertaining physic and chemistry including measurements of geomagnetic field, vectors.
3. In-situ data (C8, C9) pertaining physic and chemistry including underground/underwater instruments or sensors.
4. Reference geoids.

2.8.3 Earth Science Data User Requirements

ID	Requirements	Source
FURD-ES-1200	It is required to have access to all data, information and documents, including models and algorithms.	Interviews, SOR-1,2,3,5,6, 7, 8
FURD-ES-1210	It is required preservation of scientific data and relevant documentation for more than 20 (twenty) years.	Interviews
FURD-ES-1220	Complete cancellation of data and documents must be avoided.	Interviews
FURD-ES-1230	Time series/time windows should be greater than 20 years	Interviews
FURD-ES-1240	Documents are required describing formats and representations of data.(*) .	Interviews, SOR-1

FURD-ES-1250	Geophysics analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-1,2,3
FURD-ES-1260	Geophysics analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5
FURD-ES-1270	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-1280	Various kinds of auxiliary data are required for atmospheric and geometric correction.	Interviews, SOR-3, 6
FURD-ES-1290	Mechanism to access data and information (documents) must be simple and easy.	Interviews, SOR-1,2,3, 4
FURD-ES-1300	Mechanism to access data and information should be free of charge.	Interviews
FURD-ES-1310	Performance to access data should be at technological frontier.	Interviews, SOR-1
FURD-ES-1320	Data owner must grant integrity and quality of data, products and documents without limits.	Interviews, SOR-2, 3
FURD-ES-1330	Data access security policies must not prevent or limit accessibility.	Interviews
FURD-ES-1340	Reprocessing shall be possible on request at any time and for any timespan.	Interviews, SOR-3
FURD-ES-1350	Temporal coherence of time series must be granted or gaps must be listed clearly.	Interviews, Study
FURD-ES-1360	Data and documents concerned with campaigns must be available for more than 20 years.	Interviews
FURD-ES-1370	Geophysics analysis requires reprocessing.	Interviews, SOR-3
FURD-ES-1380	Geophysics analysis requires preservation of reprocessed data and products for 5 (five) years.	Interviews
FURD-ES-1390	Geophysics analysis requires continuous time series.	Interviews, SOR-3
FURD-ES-1400	Generated data and products must be traceable.	Interviews, SOR-3
FURD-ES-1410	Owner of C6, C7, C8 and C9 data and products should grant their availability and exploitability at any time.	Interviews, SOR-3, Study

Notes:

(*) It is to be noticed that nowadays many data and documents concerned with geophysical science domain are maintained through databases accessible via websites. You can see a typical reference in the Table 3 RTS 17. Long term maintenance of these web sites and concerned content should be considered as matter of LTDP but are not within the current scope of LTDP/FIRST. However, this point will be addressed in the frame of changes and updating to be proposed to the LTDP Guidelines and roadmap for European capabilities.

2.8.4 EO Data User Requirements

ID	Description	Source
FURD-EO-0800	Geophysics analysis requires time series of data for more than 15 (fifteen) years.	Interviews,

FURD-EO-0810	EO data and products of all categories must be available as soon as possible.	Interviews, SOR-2, SOR-3
FURD-EO-0820	Geophysics analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types.	Interviews, SOR-2, SOR-3
FURD-EO-0830	Geophysics analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5.	Interviews, SOR-2, SOR-3
FURD-EO-0840	Geophysics analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, and C3-3.	Interviews, SOR-2, SOR-3
FURD-EO-0850	Geophysics analysis requires Primary Data of type C4 of all sub-categories.	Interviews, SOR-2, SOR-3
FURD-EO-0860	Geophysics analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10	Interviews, SOR-2, SOR-3
FURD-EO-0870	Geophysics analysis requires SAR (VHR/HR/MR) from tandem/combined missions.	Interviews, SOR-2, SOR-3
FURD-EO-0880	Geophysics analysis requires OPT/Multispectral from combined missions including stereo.	Interviews, Study, SOR-2
FURD-EO-0890	Geophysics analysis requires raw and higher product (i.e. L0, L1, L2, L3).	Interviews,
FURD-EO-0900	Geophysics analysis requires secondary data: Ancillary, Auxiliary, CAL/VAL.	Interviews, SOR-1,2, 3
FURD-EO-0910	Geophysics analysis requires metadata and browse images.	Interviews, SOR-1,2, 3
FURD-EO-0920	Geophysics analysis requires Mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3
FURD-EO-0930	Geophysics analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-0940	Geophysics analysis requires EO remote sensed data coherent with campaigns, at least temporarily.	Interviews, SOR-1,2, 3
FURD-EO-0950	Geophysics analysis requires access to all available EO data free of charge.	Interviews, SOR-1,2, 3

2.8.5 Additional User data requirements

ID	Description	Source
FURD-AB-0400	Geophysics analysis requires campaign data and products (i.e. C6, C7, C8, C9) for validation of fluid dynamic models	Interviews, Study
FURD-AB-0410	Geophysics analysis requires integration with satellite remote sensed data (i.e. EO) for analysis & validation of fluid dynamic models	Interviews, Study
FURD-AB-0420	Long time series are necessary to validate models	Interviews, Study
FURD-AB-0430	Wide spectra of remote sensed data are necessary to validate models	Interviews, Study
FURD-AB-0440	Geophysics analysis requires in-situ data and products (i.e. C6, C7, C8, C9) for comparison, correction and integration to	Interviews, Study



	satellite remote sensed data (i.e. EO)	
FURD-AB-0450	Geophysics analysis requires in-situ related (i.e. C6, C7, C8, C9) documents aligned between them.	Interviews, Study
FURD-AB-0460	Geophysics analysis requires a high or very high location geo-reference and altitude measurement (i.e. x, y, z).	Interviews, Study

2.8.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C1 (SAR, all sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C3 (all sub-types)
 - i. Raw data
 - ii. Higher level
 - d. C4 (all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)
 - i. Higher level
 - f. C6 (Airborne)
 - g. C7 (Balloon)
 - h. C8 (Ground)
 - i. C9 (Hydro)
2. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries:
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
3. Metadata
4. Browse image

2.8.7 Special case

None

2.9 Earth as a system (EaaS)

2.9.1 Domain Description

This section deals with all embracing sciences concerned with Earth. The main driver is the concept that whatever is the event there are various effects on or causes from the systems in which we are used to divide Earth's science: lithosphere, hydrosphere, biosphere and atmosphere.

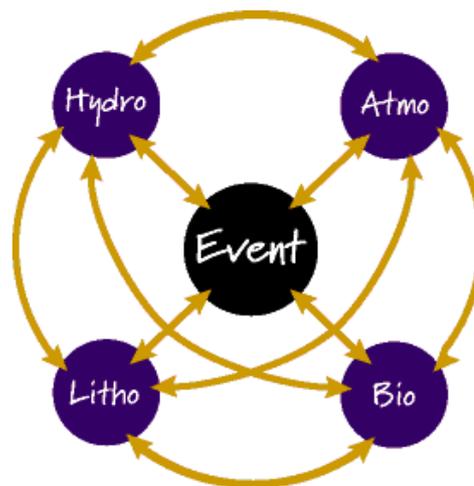


Figure 9 Sub-systems interactions (from <http://www.cotf.edu>)

We have considered and analysed requirements directly issued or retrievable from documents and relevant to the application fields:

1. **Water variability.** It is of major importance both for the functioning of aquatic species and also for the availability of water for human use: water that is only available in a few wet years cannot be considered renewable in a human time cycle. Desertification is one of the other aspects considered.
2. **Transcontinental air pollution.** It is the extension of usual pollution concept to those situations where very wide areas are involved. In principle all substances which (when present in sufficient concentration, for a sufficient time, and under certain conditions) tend to interfere with human comfort, health or welfare, and cause environmental damage. Transcontinental air pollution causes acid rain, ozone depletion, photochemical smog, and other such phenomenon. (E.g. projects: Transcontinental wildfire emissions monitoring).
3. **Shift in ecosystem structure.** Focusing to study the composite effect of climate change on different trophic levels for ecosystem structure and functioning and understand climate-induced changes from those caused by human activities and interactions.



4. **Human health and climate changes.** Analyse the impacts of changes, both negative and positive, affecting human health. (Ref. WHO <http://www.who.int/globalchange/publications/cchhsummary/en/>).
5. **Hazards assessment and mapping.** It is mainly related to the activities and projects following the resolution 44/236 of December 22nd 1989 of UN defining the International Decade for Natural Disaster Reduction (IDNDR) and successor projects/programmes as ISDR (International Strategy for Disaster).

2.9.2 General Data Needs

Scientific data and information needs are concerned with:

1. Atmosphere models and references (e.g. General Circulation Model, AGCM, OGCM, etc.).
2. Ocean models and references (e.g. General Circulation Model, Ocean Circulation Model(s), etc.).
3. Atmospheric composition data (e.g. gases, particles, etc.).
4. Ocean chemical and physical properties and composition.
5. Long-term time series for validation of models, analysis of trends and variations.

Needed types of datasets are relevant to:

1. Earth Observation space data pertaining atmospheric, land and ocean physic and chemistry.
2. In-situ data (C6, C7, C8, C9) pertaining physic and chemistry of atmosphere at different layers.
3. In-situ data (C6,C7,C8,C9) pertaining physic and chemistry of ocean and lands
4. Water and ice cycles.

2.9.3 Earth Science Data User Requirements

ID	Description	Source
FURD-ES-1500	It is required to have access to all data, information and documents, including models and algorithms.	Interviews, SOR-1,2,3,5,6,7, 8
FURD-ES-1510	It is required preservation of scientific data and relevant documentation for more than 20 (twenty) years.	Interviews
FURD-ES-1520	Complete cancellation of data and documents must be avoided.	Interviews
FURD-ES-1530	Time series/time windows should be greater than 20 years.	Interviews
FURD-ES-1540	Documents are required describing formats and representations of data.	Interviews, SOR-1

FURD-ES-1550	EaaS analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-1,2,3
FURD-ES-1560	EaaS analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5
FURD-ES-1570	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-1580	Mechanism to access data and information (documents) must be simple and easy	Interviews, SOR-1,2,3, 4
FURD-ES-1590	Mechanism to access data and information should be free of charge	Interviews
FURD-ES-1600	Performance to access data should be at technological frontier	Interviews, SOR-1
FURD-ES-1610	Data owner must grant integrity and quality of data, products and documents without limits	Interviews, SOR-2, 3
FURD-ES-1620	Data access security policies must not prevent or limit accessibility	Interviews
FURD-ES-1630	Reprocessing shall be possible on request at any time and for any timespan	Interviews, SOR-3
FURD-ES-1640	Temporal coherence of time series must be granted or gaps must be listed clearly	Interviews, Study
FURD-ES-1650	Data and documents concerned with campaigns must be available for more than 20 years	Interviews
FURD-ES-1660	EaaS analysis requires reprocessing	Interviews, SOR-3
FURD-ES-1670	EaaS analysis requires continuous time series	Interviews, SOR-3
FURD-ES-1680	Generated data and products must be traceable	Interviews, SOR-3
FURD-ES-1690	Owner of C6, C7, C8 and C9 data and products should grant their availability and exploitability at any time	Interviews, SOR-3, Study

2.9.4 EO Data User Requirements

ID	Description	Source
FURD-EO-1000	EaaS analysis requires time series of data for more than 20 (twenty) years	Interviews,
FURD-EO-1010	EO data and products of all categories must be available as soon as possible	Interviews, SOR-2, SOR-3
FURD-EO-1020	EaaS analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types	Interviews, SOR-2, SOR-3
FURD-EO-1030	EaaS analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5	Interviews, SOR-2, SOR-3
FURD-EO-1040	EaaS analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, C3-3	Interviews, SOR-2, SOR-3
FURD-EO-1050	EaaS analysis requires Primary Data of type C4 of all sub-categories	Interviews, SOR-2, SOR-3
FURD-EO-1060	EaaS analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10	Interviews, SOR-2, SOR-3
FURD-EO-1070	EaaS analysis requires SAR (VHR/HR/MR) from tandem/combined missions	Interviews, SOR-2, SOR-3
FURD-EO-1080	EaaS analysis requires OPT/Multispectral from combined missions including stereo	Interviews, Study, SOR-2
FURD-EO-1090	EaaS analysis requires raw and higher product (i.e. L0,L1,L2, L3)	Interviews,
FURD-EO-1100	EaaS analysis requires SAR /OPT/Multispectral/Other science instruments combination	Interviews
FURD-EO-1110	EaaS analysis requires metadata and browse	Interviews, SOR-1,2, 3
FURD-EO-1120	EaaS analysis requires Mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3
FURD-EO-1130	EaaS analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-1140	EaaS analysis requires EO remote sensed data coherent with campaigns, at least temporarily	Interviews, SOR-1,2, 3
FURD-EO-1150	EaaS analysis requires access to all available EO data free of charge	Interviews, SOR-1,2, 3

2.9.5 Additional User requirements

None

2.9.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data:
 - a. C1 (SAR, both sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C3 (all sub-types)



- i. Raw data
 - ii. Higher level
 - d. C4 (all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)
 - i. Higher level
 - f. C6 (Airborne)
 - g. C7 (Balloon)
 - h. C8 (Ground)
 - i. C9 (Hydro)
 - 2. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries:
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
 - 3. Metadata
 - 4. Browse image

Note: The kind of primary data of categories C6, C7, C8 and C9 are the same of the Atmosphere, Oceanography, Biosphere and Geophysics scientific domains.

2.9.7 Special case

None.

2.10 Cal/Val

2.10.1 Domain Description

This section refers to a set of best practices (i.e. engineering practices) which overall objective is to grant and improve the overall quality of sensors and instruments data and products. This section addresses mainly spaceborn sensors; however almost all needs have been confronted with in-situ instruments/sensors and airborne carried ones. These activities are relevant to:

1. **Calibration of sensors and/or instruments**; where calibration is the comparison between measurements - one of known magnitude or correctness made or set with one device and another measurement made in as similar a way as possible with a second device. The CEOS definition is: “Calibration is the process of quantitatively defining the system responses to known, controlled, signal inputs”.
2. **Validation**, by establishing objective evidences that a process (or a measurement) consistently produces a result (or a product) meeting its predefined requirements. The CEOS definition is: “Validation is the process of assessing by independent means, the quality of the data products derived from the system outputs”.

The overall goal is to increase measurement accuracy of a sensor and to identify uncertainties. In principle Cal/Val should include all sensors and measurement used in Earth Science for any kind of reason. In the current version of the document, this scope has been reduced to the CAL/VAL activities of the Earth Observation sensors (spaceborn).

Additional references with respect to Table 3 and Table 4 are:

<http://calvalportal.ceos.org/cvp/web/guest/overview>

QA4EO references

(<http://calvalportal.ceos.org/cvp/web/guest/qa4eo>)

<http://earth.esa.int/> (search for Cal/Val hints)

<http://www.esa.int/esaLP/LPearthexp.html>

<http://envisat.esa.int/calval/proceedings/>

http://earth.esa.int/pub/ESA_DOC/envisat_val_1202/proceedings/ACV/ground_based/02_kro_gnes.pdf

http://smc.cnes.fr/IASI/TN3109-Reports_CalValA_new.pdf (example)

Appl. Remote Sensing Cluster (CAF), German Aerosp. Center (DLR), Wessling, Germany (e.g. Cal/Val EnMAP)

<http://www.jpl.nasa.gov> (search for Cal/Val using the internal facility)

We have considered and analysed requirements directly issued or retrievable from public documents and relevant to the calibration and validation of the following sensors:

1. SAR (ref. C1)
2. Optical and IR (ref. C2, C3)
3. Atmosphere sensors (ref. C4)



4. Altimetry (ref. C3)
5. Gravimeters, Accelerometers (ref. C5)
6. Microwave radiometer (passive) (ref. C3)

The above described range in the list does not cover all nine data categories due to limited availability of information and documents (including interviews) concerned with all the other instruments. Interviews with Cal/Val experts and operators have been performed and some experiences from already provided support to the Calibration & Validation processes analysis have been considered too.

2.10.2 Cal/Val Engineering Data Needs

Cal/Val Engineering data and information needs are concerned with:

1. Grid cell location on fixed Earth grid (lat, lon), and Earth geoid model.
2. Characteristics of ground references for calibration (e.g. albedo, radiometry)
3. Static ancillary data (e.g. permanent masks [land, water, urban, etc.], DEM info, soil type, etc.)
4. Dynamic ancillary data (e.g. temperatures, vegetation water content, vegetation parameters, etc)
5. Weather parameters and measurements, including snow/ice
6. Atmosphere, ocean and hydrological models, land-atmosphere-ocean exchange models (including thermohaline)
7. Input data quality indicators
8. Seasonal dependencies and concerned models
9. Sun, moon and other external impacts and interferences
10. Algorithms and algorithms error budgets, algorithms testbed simulations
11. Cal/Val mission characteristics, performances and results (Cal/Val Mission)
12. Geometries and effects of viewing geometries
13. Modelling and data assimilation from other sources including other satellites (e.g. Carbon tracker)
14. Time series algorithms (see Note 1), long term time series for validation of models, analysis of trends and variations
15. Scaling effects (e.g. aircraft over heterogeneous areas)
16. Coherence among information (e.g. versions, standardisation of representations, qualification, etc)
17. Availability of documents (e.g. models, math, algorithms, missions, campaign, etc)

Note 1: As matter of fact from consulted people and documents, it seems that long time series are largely necessary and preferably on daily basis sampling. In particular, according to many interviewees the best useful sample of dataset should be daily (1 sample per day) while maintaining data for long term in order to analyze trend and variations.

2.10.3 Earth Science Data Requirements from Cal/Val Engineering

ID	Description	Source
FURD-ES-1800	It is required to have access to all data, information and	Interviews,

	documents, including models and algorithms and data concerned with reference sites (e.g. vicarious calibration).	Study, SOR-1,2,3,5,6, 8
FURD-ES-1810	Instrument and/or sensor characteristics including pre-flight or pre-operational performance measurements (e.g. spectral response, noise characteristics) must be preserved likewise sensed telemetries or transformed data (raw or higher level) .	Interviews, Study
FURD-ES-1820	Instrument and/or sensor calibration data and methods must be preserved likewise sensed telemetries or transformed data.	Interviews, Study
FURD-ES-1830	Processing algorithms and their scientific basis including complete description of any sample or mapping algorithm used in the creation of Cal/Val procedures must be preserved likewise all other datasets.	Interviews, Study
FURD-ES-1840	Complete information on any ancillary data or other dataset used in simulation or calibration of the dataset or derived product must be saved likewise all other datasets.	Interviews
FURD-ES-1850	Complete cancellation of simulation data, temporary data and documents should be avoided.	Interviews
FURD-ES-1860	Typical preservation time windows should be greater than 20 years.	Interviews
FURD-ES-1870	Documents are required describing formats and representations of data.	Interviews, SOR-1
FURD-ES-1880	Cal/Val requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-1,2,3 (SOR-1 ECV, FCDR)
FURD-ES-1890	Cal/Val analysis, including possible reprocessing, requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others for comparison purpose.	Interviews, SOR-1,2,3, 5 (SOR-1 ECV, FCDR)
FURD-ES-1900	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-1910	Several kinds of auxiliary and ancillary data are required for various corrections.	Interviews, SOR-3, 6
FURD-ES-1920	Mechanism to access data and information (documents) must be simple and easy	Interviews, SOR-1,2,3, 4
FURD-ES-1930	Mechanism to access data and information should be free of charge and open to all Cal/Val communities.	Interviews
FURD-ES-1940	Sampling of data should be on daily basis (1 sample per day) if not more frequently required.	Interviews, SOR-1
FURD-ES-1950	Reference data (e.g. independent reference) owner must grant integrity and quality of data, products and documents without limits.	Interviews, SOR-2, 3
FURD-ES-1960	Data access security policies must not prevent or limit accessibility.	Interviews
FURD-ES-1970	Cal/Val requires long time series of airborne, balloon, ground radar and in-situ data for rigorous testing of algorithms.	Interviews, SOR-3
FURD-ES-1980	Cal/Val requires long time series of airborne, balloon, ground radar and in-situ data to cover full dynamic range of instruments and algorithms.	Interviews, Study
FURD-ES-1990	Cal/Val requires long time series of airborne, balloon,	Interviews



	ground radar and in-situ data to investigate effects of changing on time series parameters.	
FURD-ES-2000	A Cal/Val activity requires partial reprocessing for testing purpose. Test results must be preserved similarly to primary data.	Interviews, SOR-3
FURD-ES-2010	Data and documents concerned with campaigns must be available all times	Interviews
FURD-ES-2020	Generated data and products must be traceable	Interviews, SOR-1,2,3,6

Note: Requirement FURD-ES-1900 is at same time required to owners of external data and provided as per produced data according to QA4EO guidelines.

2.10.4 Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C1 (SAR, all sub-types and all bands)
 - b. C2 (Optical, all sub-types)
 - c. C3 (all sub-types)
 - d. C4 (all sub-types)
 - e. C5 (all sub-types)
 - f. C6 (Airborne)
 - g. C7 (Balloon)
 - h. C8 (Ground)
 - i. C9 (Hydro)
2. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
3. Metadata
4. Browse image

2.10.5 Additional requirements

None

2.10.6 Special case

None.

2.11 Geoinformatics

2.11.1 Domain Description

Geoinformatics is the science and the technology that develops and uses information science infrastructure to address the problems of geography, geosciences and related branches of engineering. Geoinformatics has at its core the technologies supporting the processes of acquiring, analyzing and visualizing spatial data. Both geomatics and geoinformatics include and rely heavily upon the theory and practical implications of geodesy.

We have considered and analysed requirements directly issued or retrievable from public documents and relevant to the following fields:

1. Spatial based Decision Making Support Systems (DMSS).
2. Emergency Support Systems.
3. Archaeology.
4. Resource Management.
5. Human Geography.
6. Cartography.
7. Geostatistics.
8. DEM and 3D models.

Reference is made to OpenGIS (<http://www.opengeospatial.org/standards/requests/70>) and in particular to the Earth Observation Satellite Tasking Extension for OGC® Sensor Planning Service (SPS), (see document: <http://www.opengis.net/doc/IS/EOSPS/2.0>).

2.11.2 General Data Needs

Scientific data and information needs are concerned with:

1. Grid cell location on fixed Earth grid (lat, lon), and Earth geoid model, topography.
2. Precise location (GPS).
3. Spatial data on the geographical landscape, both Optical and SAR types.
4. Time series should be the longest possible.
5. Geophysical, geopolitical, and demographical data and statistics

2.11.3 Earth Science Data User Requirements

ID	Description	Source
FURD-ES-2100	It is required to have access to all data, information and documents	Interviews, Study, SOR-1,2,3,5, 6, 7, 8
FURD-ES-2110	It is required preservation of observed data and relevant documentation	Interviews, Study
FURD-ES-2120	Time series/time windows should be greater than 15	Interviews

	years	
FURD-ES-2130	Documents are required describing formats and representations of data.	Interviews, SOR-1
FURD-ES-2140	Analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, geographical data, historical data, and data generated by other research or application or available on different websites.	Interviews, SOR-1,2,3 (SOR-1 ECV, FCDR)
FURD-ES-2150	Geoinformatics analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-1,2,3, 5 (SOR-1 ECV, FCDR)
FURD-ES-2160	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-3
FURD-ES-2170	Mechanism to access data and information (documents) must be simple and easy (including websites)	Interviews, SOR-1,2,3, 4
FURD-ES-2180	Mechanism to access data and information should be free of charge	Interviews
FURD-ES-2190	Data owner must grant integrity and quality of data, products and documents without limits (including websites)	Interviews, SOR-2, 3
FURD-ES-2200	Data access security policies must not prevent or limit accessibility (including websites)	Interviews, study
FURD-ES-2210	Geoinformatics analysis requires temporal on data-time series	Interviews, SOR-3
FURD-ES-2220	Generated data and products must be traceable and provenance identified	Interviews, SOR-3

2.11.4 EO Data User Requirements

ID	Description	Source
FURD-EO-1200	Geoinformatics analysis requires time series of data for more than 20 (twenty) years	Interviews,
FURD-EO-1210	EO data and products of all categories must be available as soon as possible	Interviews, SOR-2, SOR-3
FURD-EO-1220	Geoinformatics analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2 of all bands and types	Interviews, SOR-2, SOR-3
FURD-EO-1230	Geoinformatics analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5	Interviews, SOR-2, SOR-3
FURD-EO-1240	Geoinformatics analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10	Interviews, SOR-2, SOR-3
FURD-EO-1250	Geoinformatics analysis requires SAR (VHR/HR/MR) from tandem/combined missions	Interviews, SOR-2, SOR-3
FURD-EO-1260	Geoinformatics analysis requires OPT/Multispectral from combined missions including stereo	Interviews, Study, SOR-2
FURD-EO-1270	Geoinformatics analysis requires raw and higher product	Interviews,



	(i.e. L0, L1, L2, L3)	
FURD-EO-1280	Geoinformatics analysis requires SAR /OPT/Multispectral/Other science instruments combination	Interviews
FURD-EO-1290	Geoinformatics analysis requires metadata and browse	Interviews, SOR-1,2, 3
FURD-EO-1300	Geoinformatics analysis requires Mission related documentation aligned with data and product versions. Examples of documents are: documents concerned with platform performances, instruments performances, known errors, quality trends, etc.	Interviews, SOR-1,2, 3
FURD-EO-1310	Geoinformatics analysis requires description of quality and uncertainty of measures, data and products.	Interviews, SOR-1,2, 3
FURD-EO-1320	Geoinformatics analysis requires access to all available EO data free of charge (possibly)	Interviews, SOR-1,2, 3

2.11.5 Additional User requirements

None

2.11.6 Summary Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C1 (SAR, all sub-types and all bands)
 - i. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C5 (all sub-types)
 - i. Higher level
2. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries:
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation
3. Metadata
4. Browse image

2.11.7 Special case

None

2.12 GMES

This section deals with GMES context and is aimed at explaining in more detail the indirect capture of LTDP User Requirements from a specific service oriented context.

GMES (Global Monitoring for Environment and Security) is the European Union lead Initiative for the establishment of a European capacity for Earth Observation (see also <http://www.gmes.info/>). Currently the following R&D projects have been funded under the 7th Framework Programme (FP7) of the European Union:

- 1) Geoland2 (land monitoring)
- 2) MyOcean (marine monitoring)
- 3) MACC (atmosphere monitoring)
- 4) SAFER (emergency response)
- 5) G-MOSAIC (security)

Requirements impacting LTDP of Earth Science data from the above listed services, that constitute the Core GMES operational services, have been identified through the questionnaires and through the analysis of documents in Table 3 - Reference documents for requirements
SOR-2.

2.12.1 General Data Needs

Scientific data and information needs are concerned with:

1. Atmosphere and ocean models and references (e.g. General Circulation Model, AGCM, OGCM, etc.).
2. Atmospheric composition data (e.g. gases, particles, etc.).
3. Ocean physics and chemical static/dynamic data.
4. Land physics and chemical data.
5. Hydro-geologic models and data.
6. Bio-physical parameters.
7. Geography and topography.
8. Land surface models.
9. Retrieval algorithms.
10. Static/dynamic thematic masks (e.g. sea, land, hydro, urban, risks, etc.) .
11. Earth Observation space data (C1, C2, C3, C4 and C5) both Primary and Secondary.
12. In-situ data (C6, C7, C8, C9) pertaining physic and chemistry.
13. Long term time series for analysis of trends and variations.
14. Time series is variable and up to 20 years.
15. On-time availability of data (i.e. real-time, near-real-time) and high volume throughput.
16. Coherence among information (e.g. interoperability)
17. Availability of documents (e.g. models, math, algorithms, missions, campaign, etc)

2.12.2 Earth Science Data User Requirements from GMES Services

ID	Description	Source
FURD-ES-2400	It is required to have access to all data, information and documents	Interviews, SOR-2, Study
FURD-ES-2410	Complete cancellation of data and documents must be avoided	Interviews, SOR-2, Study
FURD-ES-2420	Documents are required describing formats and representations of data.	Interviews, SOR-2, Study
FURD-ES-2430	GMES service analysis requires possibility to combine data from various sources including but not limited to: EO satellite based sensors, in-situ sampled data, and historical data, data generated by other scientific research or application.	Interviews, SOR-2, Study
FURD-ES-2440	GMES service analysis requires all possible EO data without limits to types of sensors, bands, resolution, measurement cycles, or others.	Interviews, SOR-2, Study
FURD-ES-2450	Indicators of quality/reliability of data and products, including auxiliary data and products (e.g. precision, tolerance, known errors, limits, etc.) are required.	Interviews, SOR-2, Study
FURD-ES-2460	Various kinds of auxiliary data are required for atmospheric and geometric correction.	Interviews, SOR-2, Study
FURD-ES-2470	Mechanism to access data and information (documents) must be simple and easy (ref. DAP)	Interviews, SOR-2, Study
FURD-ES-2480	Data owner must grant integrity and quality of data, products and documents without limits	Interviews, SOR-2, Study
FURD-ES-2490	Data access security policies must not prevent or limit accessibility	Interviews, SOR-2, Study

2.12.3 Example: GMES Services Dataset Composition

The Data set composition related to the needs of the GMES services has been derived taking into account the data necessary to implement each specific service. GMES services, being operational services, are addressed to non-scientific users (e.g. decision makers). GMES services providers are requiring to Earth Science data providers, the data and information necessary and useful to provide their services. GMES Services relies on the other hand on the previous work performed by researchers to create models for analysis and forecasting needed by the services themselves. Creation and maintenance of those models requires the preservation of the relevant EO and in-situ data as well as models for prediction or forecast of events. Thus GMES is both a direct (from the services) and an indirect (from the models used in the services) source of LTDP requirements.

Typical GMES services datasets include:

- 1) European Soil Database: it includes various kinds of datasets concerned with different services and applications (e.g. organic, metals, carbon cycle, etc, see http://eusoils.jrc.ec.europa.eu/library/esdac/esdac_access2.cfm) and metadata (see http://eusoils.jrc.ec.europa.eu/library/esdac/Esdac_DetailData2.cfm?id=1)



- 2) MyOcean dataset includes: Numerical models (MFC) global and regional, in-situ and remote sensing Observations (TAC) (ref. <http://www.myocean.eu.org/>).
- 3) MACC dataset looks at:
 - a) Aerosol data retrievals from the (A)ATSR, AVHRR, GOME, GOME-2, SCIAMACHY and SEVIRI instruments.
 - b) Carbon-dioxide and methane retrievals from the AIRS and IASI instruments.
 - c) Ozone and nitrogen dioxide retrievals from OMI, SCIAMACHY, GOME and GOME-2.
 - d) An updated inventory of global and European-regional emissions (maps).
 - e) Near-real-time and retrospective analyses of fire emissions.
 - f) Regional air-quality analysis and forecasting systems.
- 4) SAFER dataset is based on:
 - a) EO remote sensed data (C1, C2, C3, C4 and C5)
 - b) In-situ data (currently are prevailing of type C8/C9)
 - c) Geography, topography, DEM
 - d) Asset maps, resources and inventories
 - e) Statistics at regional, local and municipality level (e.g. population, houses...)
- 5) G-Mosaic deals mainly with (*):
 - a) VHR SAR (C1) and optical stereo data (C2)
 - b) DEMs and objects reconstruction algorithms
 - c) HR SAR (C1) data as supporting or second chance

The following items should also be considered part of the dataset:

- 1) The global ECVs and their assimilation into ocean/ice/land surface models.
- 2) Local ocean/ice/land cover zooming on ‘hot spot areas’ (e.g. NATURA2000, coastal areas).

Notes

(*) G.Mosaic service for the security community deals with ad-hoc requests for SAR-VHR and SAR-HR data including data available outside Europe and not considered in the scope of this study.

2.12.4 EO Data Requirements from GMES Services

ID	Description	Source
FURD-EO-1400	GMES service analysis requires time series of data up to archive available data	GMES, DAP
FURD-EO-1410	EO data and products of all categories must be available as soon as possible (i.e. real time)	GMES, DAP
FURD-EO-1420	GMES service analysis requires Primary Data of type C1 both sub-category C1-1 and C1-2	GMES, DAP

FURD-EO-1430	GMES service analysis requires Primary Data of type C2 of sub-categories C2-1, C2-2, C2-3, C2-4 and C2-5	GMES, DAP
FURD-EO-1440	GMES service analysis requires Primary Data of type C3 of sub-categories C3-1, C3-2, C3-3	GMES, DAP
FURD-EO-1450	GMES service analysis requires Primary Data of type C4 of all sub-categories	GMES, DAP
FURD-EO-1460	GMES service analysis requires Primary Data of type C5 of sub-categories C5-1, C5-2, C5-3, C5-4, C5-6, C5-9, C5-10	GMES, DAP
FURD-EO-1470	GMES service analysis requires SAR (VHR/HR/MR) from tandem/combined missions	GMES, DAP
FURD-EO-1480	GMES service analysis requires OPT/Multispectral from combined missions including stereo	GMES, DAP
FURD-EO-1490	GMES service analysis requires raw and higher product (i.e. L0, L1, L2, L3)	GMES, DAP
FURD-EO-1500	GMES service analysis requires metadata and browse	GMES, DAP
FURD-EO-1510	GMES service analysis requires description of quality and uncertainty of measures, data and products.	GMES, DAP
FURD-EO-1520	GMES service analysis requires access to all available EO data	GMES, DAP

2.12.5 Additional requirements

None

2.12.6 Earth Science Dataset Composition

Typical dataset includes:

1. Primary data :
 - a. C1 (SAR, all sub-types and all bands)
 - i. Raw data
 - ii. Higher level
 - b. C2 (Optical, all sub-types)
 - i. Higher level
 - c. C3 (Radiometric, all sub-types)
 - i. Raw data
 - ii. Higher level
 - d. C4 (Atmosphere Chemistry, all sub-types)
 - i. Higher level
 - e. C5 (all sub-types)
 - i. Higher level
 - f. C6 (Airborne) TBC depending by the evolution of services**
 - g. C7 (Balloon) TBC depending by the evolution of services**
2. Secondary data
 - a. Ancillary data
 - b. Auxiliary data to process telemetries:
 - c. Cal/Val data including calibration documentation and history
 - d. Mission related documentation



3. Metadata
4. Browse images

2.12.7 Special case

None.

2.13 Climate Change Initiative (CCI)

The Climate Change Initiative is the ESA initiative in response to the UNFCCC and IPCC (ref. post Kyoto assessments and initiatives).

Climate change is arguably the greatest challenge facing mankind in the twenty-first century. Its importance has been recognised in recent reports from the IPCC and from UNFCCC, and the overwhelming economic consequences are set out in the so called Stern Report (available at http://ff.org/centers/csspp/pdf/20061104_stern.pdf, critical analysis at <http://www.iied.org/pubs/pdfs/11501IIED.pdf>, impact model example at http://unfccc.int/adaptation/nairobi_work_programme/knowledge_resources_and_publications/items/5447.php).

Here are reported and referenced data preservation needs mainly addressing ECV Essential Climate Variables (reference is to Table 3 SOR-6). CCI working groups and a Climate Modelling User Group (CMUG) have identified requirements concerned with the data needed to generate the ECVs in the framework of the ESA CCI initiative (CCI EO Data Requirements, ESA Peter Regner, issue September 2010. This document is referenced as source of requirements for LTDP/FIRST requirements in Table 3 SOR-3. CCI working groups are currently in progress reviewing ECVs concepts and detailed definitions.

The Global Climate Observing System (GCOS) requirements are on the other hand taken from the "Implementation Plan for the Global Observing system for climate in support of the UNFCCC, August 2010" document <http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf> (see reference in Table 3 SOR-1).

2.13.1 Climate Change Initiative Data Needs

Data and information needs are concerned with:

1. Atmosphere and ocean models and references (e.g. General Circulation Model, AGCM, OGCM, etc.)
2. Atmospheric composition data (e.g. gases, particles, etc.)
3. Ocean composition and dynamic
4. Earth Observation space data pertaining atmospheric physic and chemistry of atmosphere (C1, C2, C3, C4 and C5) both Primary and Secondary
5. In-situ data (C6,C7) pertaining physic and chemistry of atmosphere at different layers
6. In-situ data (C8,C9) pertaining physic and chemistry of atmosphere at low layers of interface with land and water
7. Long term time series for validation of models, analysis of trends and variations.
8. Time series should be for more than 20 years
9. Coherence among information (e.g. versions, standardisation of representations, qualification, etc)
10. Availability of documents (e.g. models, math, algorithms, missions, campaign, etc)

Earth Science Data User Requirements coming from the Climate Change Initiative (CCI) and from GCOS are spread in the tables "Earth Science Data User Requirements" of each Earth Science application domain described in the previous sections and are identified in the Source column by the source of reference SOR-3, SOR-6 and SOR-1.

3. SUMMARY OF REQUIREMENTS AND NEEDS

Broadly speaking, the capture of user requirements and needs concerned with the long term preservation of Earth Science data and with the European LTDP Common Guidelines and themes has confirmed some basic assumptions, e.g. in terms of what data should be preserved in the long term and has contributed to fill gaps and to raise interesting aspects related to Earth Science data preservation not considered before.

Data preservation of Earth Science data is considered strategic for the future of any kind of research and development of practical applications in this domain and is a general expectation by users. The requirements survey has allowed to have a more precise and consolidated dimension of what (i.e. datasets) need to be preserved as well as a greater confidence and confirmation about the duration of time series needed for Earth Science applications.

Security aspects for data access and data exploitability are on the other hand not the main priority for users. Security aspects are considered fully in charge of data owners while from the user point of view this aspect is considered as a sort of toll to gain access to data and information. Data exploitability is a difficult point to cover and has a wide spectrum of answers. Users assume data are exploitable and further exploitability is part of their job and domain of application. According to user perception and expectations the data-owner simply should make data available according to suitable systems and access methods without addressing exploitability. This aspect is on the other hand very critical if we think about future possible uses of the data that might also be not known today. This requires that data providers preserve today not only the data and information needed for today exploitability of the data but also any other information and knowledge associated to the data and that might be needed in future for other possible uses of the data themselves. The definition of this “data associated knowledge” is a critical aspect to be addressed and consolidated today to guarantee at the maximum extent that data are exploitable also by future users.

Some points of interests in addition to what contained in the European LTDP Common Guidelines are represented by:

- a) The necessity to have coherence among all the components of any dataset. This means that the entire group of data must be consistent and logically ordered.
- b) The necessity to have quality indicators of data and information in order to ensure trustiness of what will be used for other application.
- c) The necessity to have traceability of data and information allowing a back propagation to the sources of different data and information (either human or machine generated).
- d) The necessity to have continuity in series of data or information from similar instruments (e.g. same band SAR) or the exigency to reduce as much as possible gaps in sequential series of information (or data).



- e) Homogeneity of preservation, that means the necessity to preserve groups of information (i.e. datasets) in such a way that the whole is retrievable and usable later on. This is not currently guaranteed as datasets are often settled with components coming from different owners having different technologies and different utilisation objectives.
- f) Wasting of poor quality optical EO data should be more carefully analysed being possible utilisation by other entities. One known example is concerned with optical products with high cloudiness index. In this case, products could be used by communities interested in cloud related researches (e.g. clouds and aggregation of aerosols).

3.1 Summary of needs and requirements from ES user community

The synthetic list of requirements is:

1. ES data preservation is necessary to create, maintain and allow evolution of the understanding of scientific events.
2. ES data preservation is a common and widely recognized need to allow performing today and future Earth Science research and application activities.
3. ES data preservation allows future re-use and paves the way to further knowledge growth.
4. Preserve all data forever is the mantra for the scientific users.
5. A user centred vision should be applied rather than a provider centred one.
6. Availability and accessibility to all data and documents should be open and free and access limitations/rights are a concern for users.
7. Harmonization among different actors and data providers should be a primary objective in order to ensure an easy access to all scientific data.
8. All elements necessary to have the complete knowledge on a specific data object need to be preserved in addition to the data itself. Preservation of all documents from inception till operations of any scientific instrument or sensor is required and mandatory to preserve the knowledge.
9. All data and derived products should have associated a Quality Indicator (QI) based on documented quantitative assessment of its traceability to community agreed (ideally SI) reference standards. Reliability of scientific data to be used should be addressed.
10. Time-series durations/windows needs are different for different applications and utilizations; however the long-term preservation is considered necessary (for more than 15 years for half of the interviewed users).



11. Periodic reprocessing of data is a necessity and the typical period where the two-versions (previous-current) should be preserved is around 5 (five) years.
12. Access to data and services should be simple and free of charge.
13. Security access issues are matter of providers/owners and should minimize the effort on user side.



4. SUMMARY OF EARTH SCIENCE DATA CATEGORIES AND COMPOSITION OF THE RELATED DATASET TO BE PRESERVED

4.1 Earth Observation (space-born) data needs

Categories of data have been identified in section #1.6 Data Categories and classification. The Earth Observation data categories needed in the different application fields and international initiatives analysed in the study are the following:

1. **C1:** SAR imaging missions/sensors, high and very high resolution (different radar bands).
2. **C2:** Multi-spectral imaging missions/sensors, high and very high resolution.
3. **C3:** Medium resolution Land and Ocean monitoring missions/sensors (e.g. wide swath ocean colour and surface temperature sensors, altimeter, etc).
4. **C4:** Atmospheric chemistry missions/sensors.
5. **C5:** Other Scientific missions/sensors.

As matter of fact, ordering statistics as well as interviews and questionnaires are converging to say that all the available data are needed by the scientific community as well as for services and projects of various kinds. According to the study results all these data should be preserved in the long term as time series spanning 20 years and more (37% of users) or 15 years and more (46% of users) are needed for several different applications.

As a conclusion we can therefore state that:

1. For Earth Science related research and scientific applications, all available Earth Observation data and products in all the available frequency bands, resolutions or with other characteristics are needed. Earth Science adapts its methods and way-of-doing caring of available data and products. Researchers address their requests to the providers having the most suitable data for their needs. Combined missions (i.e. SAR Tandem, SAR/Optical) are also of great interest.
2. For operational Services (e.g. GMES) all available Earth Observation data and products are necessary as for Earth Science research applications. However in several cases higher performances (not addressed in this document) are needed in terms of data access timeliness and resolution of data.

4.2 EO dataset composition

The composition of the EO dataset for data categories C1-C5 and for all type of instruments and bands to be preserved in the long term according to the requirements collected in the study is the following:

- a) Primary Data:
 - i. Raw data and Level 0 data or at least Level 0 data plus algorithms and procedures for their transformation.
 - ii. Global or higher level (e.g. Level1, Level2) mission products when systematically generated as part of the mission requirements and/or reprocessed.
- b) Secondary data:
 - i. Ancillary data.

- ii. Auxiliary data.
- iii. CAL/VAL databases.
- iv. Mission related documentation.
- c) Metadata
- d) Browse images when generated

For each of the data listed in points a) and b) also the following information has to be generated and preserved: Representation Information (structure, semantic and other representation information), Packaging Information, Preservation Descriptive Information (Reference, Provenance, Context, Fixity and Access Conditions/Rights). This information is needed to allow understandability and usability of the data.

The different composing elements of the dataset are defined in more detail below:

Data Levels:

- **Raw Data:** The physical telemetry payload data as received from the satellite (and recorded at the stations or received via other parties), i.e. a serial data stream without de-multiplexing. These data are not computer compatible.
- **Level 0:** Reconstructed unprocessed data at full space-time resolution with all available supplemental information to be used in subsequent processing (e.g. ephemeris, health and safety) appended.
- **Level 1A:** Reconstructed unprocessed data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and geo-referencing parameters (e.g. ephemeris) computed and appended but not applied to the Level 0 data.
- **Level 1B:** Radiometrically corrected and calibrated data in physical units at full instrument resolution as acquired.
- **Level 1C:** L1B data orthorectified, re-sampled to a specified grid.
- **Level 2:** Derived geophysical parameters (e.g. sea surface temperature, leaf area index) at the same resolution and location as Level 1 source data.
- **Level 3:** Data or retrieved geophysical parameters which have been spatially and/or temporally re-sampled (i.e. derived from Level 1 or 2 products), usually with some completeness and consistency. Such re-sampling may include averaging and compositing.
- **Level 4:** Outputs or results from models using lower level data as inputs and, thus, not directly derived from the instruments.

Ancillary data

Ancillary data can be broadly defined as those used to determine when (and how) an instrument was acquiring data, where an instrument was located, where an instrument was



pointed and what it was targeting (e.g. a surface or atmospheric feature), how those targets would appear at the time of observation and what else of potential significance to science data analysis was occurring. Actual list and content of ancillary files are strongly dependent on the mission and satellite characteristics and on the on-board sensors. A non-exhaustive list of ancillary data for a space-born instrument is reported below:

1. Satellite ephemeris
2. Satellite navigation, attitude data and orbital state vector
3. Satellite orbit (predicted, precise, etc)
4. Observation counts, orbit counts
5. On-board counters/timers
6. Calibration temperature counts
7. Antenna/Sensor noise temperature coefficients
8. Time (e.g. time of acquisition, UTC, others)
9. Latitude (e.g. relative to the acquisition)
10. Longitude (e.g. relative to the acquisition)
11. Sun azimuth (e.g. with respect to the orbit)
12. Sun elevation (e.g. with respect to the orbit)
13. Earth incidence
14. Earth azimuth
15. Data quality (e.g. BER)
16. GPS-UTC conversion
17. Files for constants and/or known errors

Auxiliary data

Auxiliary data are all data files used to generate a product, other than the direct measurements of the instrument. Auxiliary data include calibration data measured on-board but not part of the main measurements of the instrument, external calibration files from sources other than the satellite, processor configuration files, and any other files needed by instrument processors. Actual list and content of auxiliary files are strongly dependent on the mission and satellite characteristics and on the on-board sensors. A non-exhaustive list of auxiliary data for a space-born instrument is reported below:

1. AOCS corrections
2. Solar radiance
3. Instrument housekeeping table
4. Instrument non linearity correction
5. Scaling correction
6. Algorithm correction (for specific cases)
7. Gridding/re-gridding tolerance
8. Atmospheric correction factors
9. Temperature correction factors
10. Mid-long term drift correction factors
11. Failure correction factors

CAL/VAL databases

These databases contain all the information related to the calibration and validation of the instrument. Actual content is strongly dependent on the mission and satellite characteristics



and on the on-board sensors. A non-exhaustive list of the content of such a database for a space-born instrument is reported below:

1. Calibration and validation data
2. Processing/reference validation data sets
3. History of calibration/validation and qualification process used and data reliability (e.g. calibration methods, uncertainty/errors).
4. Reference sites databases and vicarious calibration references

Mission related documentation

Please, see section 1.6.1 Documentation List for details.

Metadata

Metadata is intended as information describing significant aspects of a resource (Earth Observation space data in this context). They are created for the purposes of data search, discovery and access management and may exist at various levels, typically from data collection through to the individual variables of each data file in a collection.

Browse images

Browse images are created (not for all instruments) for the purposes of data search and discovery and provide a browse of the final product to allow users to evaluate their suitability for their needs.

4.3 Earth Science data needs

Furthermore, it is to be noted that also all data pertaining to categories C6, C7, C8 and C9 must be preserved. These types are concerned with in-situ data and products as well as with campaigns. Campaigns can be considered from two different aspects. On one side campaigns are organised for Cal/Val activities concerned with measurement performances (both ground and satellite), validation of instruments and measures and similar activities. On the other side, campaigns are used for a specific research thus consider utilisation of different methods for analysis (e.g. satellite, airplanes, balloons, ground sensors, ocean sensors, others).

For Earth Science all available in-situ data of all categories from any instrument, in any band, resolution and performances are needed. Earth Science adapts its methods and way-of-doing caring of available data and products. Researchers address their requests to the providers having the most suitable data for their needs. Typically, data are available via web-sites (databases interfaced via web sites) or delivered via subscriptions. In-situ data are used together with other data (e.g. EO data, remote sensed) or standing alone depending by needs and interest of the specific research or investigation.

4.4 Earth Science dataset composition

The composition of the dataset for data categories C6-C9 and for all type of instruments to be preserved in the long term according to the requirements collected in the study is the following:

- a) Primary Data:
 - i. Raw data (instantaneous non interpolated data as acquired by the sensor/platform and recorded at the stations or received via other parties). Data can be publicly available via software interfaces (databases, network interfaces) and direct access to data loggers is restricted to the owners.
 - ii. Interpolated or processed data when systematically generated as part of the mission requirements and/or reprocessed.
- b) Secondary data:
 - i. Ancillary data.
 - ii. Auxiliary data.
 - iii. CAL/VAL databases.
 - iv. Mission related documentation.
- c) Metadata (if any)
- d) Browse images when generated

For each of the data listed in points a) and b) also the following information has to be generated and preserved: Representation Information (structure, semantic and other representation information), Packaging Information, Preservation Descriptive Information (Reference, Provenance, Context, Fixity and Access Conditions/Rights). This information is needed to allow understandability and usability of the data.

The different composing elements of the dataset are defined in more detail below:

Data Levels:

Raw Data: The physical telemetry payload data as received from the data logger (and recorded at the stations/logger or received via other parties). Generally, these data are already computer compatible.

Interpolated/Processed data: typically one processing level for acquisition correction is applied (e.g. BER reduction, scaling factors application, average calculation, others). Further processing levels depend by the actual utilisation of measures and interpolation necessities (e.g. long term average, long term trends).

Ancillary data

Ancillary data can be broadly defined as those used to determine when (and how) an instrument was acquiring data, where an instrument was located, if is a single element or belonging to a network of instruments/sensors. A non-exhaustive list of ancillary data for in-situ instruments is reported below:

1. Geoid, and georeference
2. GPS data
3. Digital Elevation Model (DEM)
4. Sampling and sequencing counters



5. Scaling factors
6. Platform dependent (e.g. glide profile, pressure profile, Environmental Sample Processor ESP, etc.)

Auxiliary data

Auxiliary data are all data files used to generate a product, other than the direct measurements of the instrument. Actual list and content of auxiliary files are strongly dependent by the type of instrument and sensor and by stand-alone instrument or by the type of network of instruments/sensors. A non-exhaustive list of auxiliary data for a space-born instrument is reported below:

1. Single/multiple element and profiling
2. Attitude, including roll/pitch/heading
3. Motion parameters (in case of not-fixed instruments)
4. Ambient state
5. Weather parameters
6. Instrument housekeeping table
7. Instrument non linearity correction
8. Scaling correction
9. Algorithm correction (for specific cases)
10. Atmospheric correction factors
11. Temperature correction factors and thermal profile
12. Local correction factors
13. Mid-long term drift correction factors
14. Failure/recovery correction factors

CAL/VAL databases

These databases contain all the information related to the calibration and validation of the instrument/sensor. Actual content is strongly dependent type of instrument, sensor and eventual network of instrument/sensor (e.g. buoys' network, boats fleet, seismic sensor network). A non-exhaustive list of the content of such a database for in-situ instruments is reported below:

1. Calibration and validation data
2. Processing/reference validation data sets
3. History of calibration/validation and qualification process used and data reliability (e.g. calibration methods, uncertainty/errors).

Mission related documentation

Please, see section 1.6.1 Documentation List for details.

Furthermore, instrument/sensor technical data sheets are considered part of the instrument documentation.



Metadata

Metadata is intended as information describing significant aspects of resource (instruments and network of instruments in this context). They are created for the purposes of data search, discovery and access management and may exist at various levels, typically from data collection through to the individual variables of each data file in a collection. Frequently are available through websites (e.g. <http://www.coriolis.eu.org/Observing-the-ocean/Observing-system-networks/GOSUD>)

Browse images

Browse images are created (not for all instruments) for the purposes of data search and discovery and provide a browse of the final product to allow users to evaluate their suitability for their needs (e.g. <http://www.coriolis.eu.org/Data-Services-Products/View-Download/Browse-T-S-maps>)