ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

System Quality Assurance Document

Atmospheric Composition ECVs

Issue 2, Version 1.0

Issued by: DLR / Thomas Popp & Oleg Goussev

Date: 31/08/2023

Ref: C3S2_312a_Lot2-D-WP3-SQAD-v2_202208_v1.0

Official reference number service contract: 2021/C3S2_312a_Lot2_DLR/SC1

PROGRAMME OF THE EUROPEAN UNION

IMPLEMENTED BY ECMWF

Contributors

DLR
Oleg Goussev
Thomas Popp

History of modifications

Version	Date	Description of modification	Chapters / Sections
C3S			
Issue 1			
1.0	29.04.2019	First issue	
1.1	21.05.2019	Minor corrections as suggested by ASSIMILA	
Issue 2			
	30.09.2019	Re-issued without change	
Issue 3			
	03.04.2020	Section 1 updated: complete CDS access structure	
Issue 4			
	03.10.2020	one minor correction (start of section 3)	
Issue 5			
	31.03.2021	Minor updates to status 03/2021	
C3S2			
Issue 1			
	19.08.2022	Minor updates for C3S2	
V1.1	17.11.2022	Minor corrections	Template updated
V1.2	12.12.2022	Editorial corrections	
Issue 2			
V1.0	31.08.2023	Minor updates for year2	

List of datasets covered by this document

This main System Quality Assurance Document covers all datasets of Lot2 “Atmospheric Composition ECVs”. Detailed lists of those datasets specific for the three ECVs included (ozone, aerosols, greenhouse gases) are provided in Annexes A – C.

Related documents

Reference ID	Document
D1	Annex A to this document: SQAD for ozone products
D2	Annex B to this document: SQAD for aerosol products
D3	Annex C to this document: SQAD for greenhouse gas products

Acronyms

We list here the general acronyms applicable to all three ECVs of this Lot; all ECV-specific acronyms are introduced in the three specific Annexes.

Acronym	Definition
BIRA-IASB	Royal Belgian Institute for Space Aeronomy
C3S	Copernicus Climate Change Service
CAMS	Copernicus Atmosphere Monitoring Service
CCI	Climate Change Initiative (ESA)
CDR	Climate Data Record
CDS	Climate Data Store
CEOS	Committee on Earth Observation Satellites
CMS	Content Management System
CPU	Central Processing Unit
CSV	Comma-Separated Values format
CUS	Copernicus User Support
DFD	German Remote Sensing Data Centre (DLR)
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EOS	Earth Observing System
EQC	Evaluation and Quality Control
ERA	ECMWF Re-Analysis
ESA	European Space Agency
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAQ	Frequently Asked Question
FP5	Fifth Framework Programme (EU)
FP6	Sixth Framework Programme (EU)
FP7	Seventh Framework Programme (EU)
FTP	File Transfer Protocol

GCOS	Global Climate Observing System
GHG	Greenhouse Gas(es)
H2020	Horizon 2020 (EU research and development programme)
HPC	High Performance Computing/Computer
HQ	Headquarters
IASB-BIRA	Royal Belgian Institute for Space Aeronomy
ICDR	Intermediate Climate Data Record
IUP-UB	Institute of Environmental Physics, University of Bremen
L1	Satellite Data Processing Level 1
L2	Satellite Data Processing Level 2
L3	Satellite Data Processing Level 3
L4	Satellite Data Processing Level 4
Lille	CNRS-Lille (Lille Observatoire Atmospherique LOA, ICARE data center)
NASA	National Aeronautics and Space Administration (USA)
NetCDF	Network Common Data Form (data file format)
OPeNDAP	Open-source Project for a Network Data Access Protocol
PI	Principal Investigator
PROMOTE	PROtocol MOniToring for the GMES Service Element: Atmosphere
PUGS	Product User Guide and Specification
R&D	Research and Development
RAM	Random Access Memory
RCS	Revision Control System (version management software)
SQAD	System Quality Assurance Document
SVN	Subversion (version management software)
TCDR	Thematic Climate Data Record
UB	University of Bremen
UNFCCC	United Nations Framework Convention on Climate Change
WAN	Wide Area Network
WMO	World Meteorological Organization
WP	Work Package

General definitions

Essential climate variable (ECV)
An ECV is a physical, chemical, or biological variable or a group of linked variables that critically contributes to the characterization of Earth's climate *(Bojinski et al., 2014)*.

Climate data record (CDR)
The US National Research Council (NRC) defines a CDR as a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change *(National Research Council, 2004)*.

Fundamental climate data record (FCDR)
A fundamental climate data record (FCDR) is a CDR of calibrated and quality-controlled data designed to allow the generation of homogeneous products that are accurate and stable enough for climate monitoring.

Thematic climate data record (TCDR)
A thematic climate data record (TCDR) is a long time series of an essential climate variable (ECV) *(Werscheck, 2015)*.

Intermediate climate data record (ICDR)
An intermediate climate data record (ICDR) is a TCDR which undergoes regular and consistent updates *(Werscheck, 2015)*, for example because it is being generated by a satellite sensor in operation.

Satellite data processing levels
The NASA Earth Observing System (EOS) distinguishes six processing levels of satellite data, ranging from Level 0 (L0) to Level 4 (L4) as follows *(Parkinson et al., 2006)*.

L0	Unprocessed instrument data
L1A	Unprocessed instrument data alongside ancillary information
L1B	Data processed to sensor units (geo-located calibrated spectral radiance and solar irradiance)
L2	Derived geophysical variables (e.g. ozone column) over one orbit
L3	Geophysical variables averaged in time and mapped on a global longitude/latitude horizontal grid
L4	Model output derived by assimilation of observations, or variables derived from multiple measurements (or both)

Manifest file
A manifest file is a complete list of all individual datasets contained in an interface; for each datasets the full link where it can be accessed from external is provided.

Gap filling
When a combination of an existing CDR and one or more ICDR extensions of it exhibit gaps or inconsistencies, the missing data or inconsistent data pieces are processed to provide a completed output data record version.

Reprocessing
When a validated new algorithm version or a new version of an input dataset is available, a complete data record from one sensor is processed to provide a new output data record version.

Level2 user support activities
While the central C3S user support operated by ECMWF responds to all general user queries, any specific queries which need involvement of the experts are elevated to “Level2 user support” which is then handed over for response to the responsible contract (e.g. in Lot2).

Table of Contents

Scope of the document

This document provides an overview of the processing system for atmospheric composition ECVs within the Copernicus Climate Change contract C3S2_312a_Lot2 (11/2021 – 4/2024), which covers processing, quality assurance and user support for Climate Data records (CDRs) of three Essential Climate Variables: Ozone (O3), Aerosol properties (AER), Greenhouse Gases (GHG: covering CO_2 and CH_4). Since this system is highly decentralised, this document covers only the system quality control of its central elements (data interface to the Climate Data Store, general principles for system procedures, user support). All other aspects on decentralised levels are described in three distinct Annexes for the O3, AER and GHG systems and their underlying decentralised processing nodes.

Executive summary

The architecture of C3S2_312a_Lot2 is highly decentralised on two levels: Firstly, based on scientific expertise as well as historical development and specific differences in the required input datasets and the applied retrieval algorithms, the processing is organized in three sub systems for the three ECVs (O3, AER, GHG). Secondly, within each sub system several different datasets are processed at a number of decentralised service provider institutions. These respond to the needs for a set of variables to be provided for each ECV, exploiting measurements from different appropriate instruments and applying different suitable algorithms.

With its decentralised system, the contract C3S2_312a_Lot2 provides a satellite-based set of complementary ECV data records with full traceability of all input and auxiliary datasets and algorithm versions. The consortium consists of members of the predecessor ESA CCI and C3S_312a, C3S_312b consortia which allows continuity of the long-term CDR dataset processing tasks within the context of C3S and maintains the close link / overlap with algorithm development in CCI.

While processing with the associated system maintenance and upgrading at each decentralised processing node and the included product quality assurance is the responsibility of each individual service provider, a common quality control is coordinated for each ECV (O3, AER, GHG). Furthermore, collecting the output data and technical format control is also conducted for each of the three ECVs. Finally, all datasets are pushed to DLR as central node which serves all output data files to the climate data store (CDS).

1. System overview

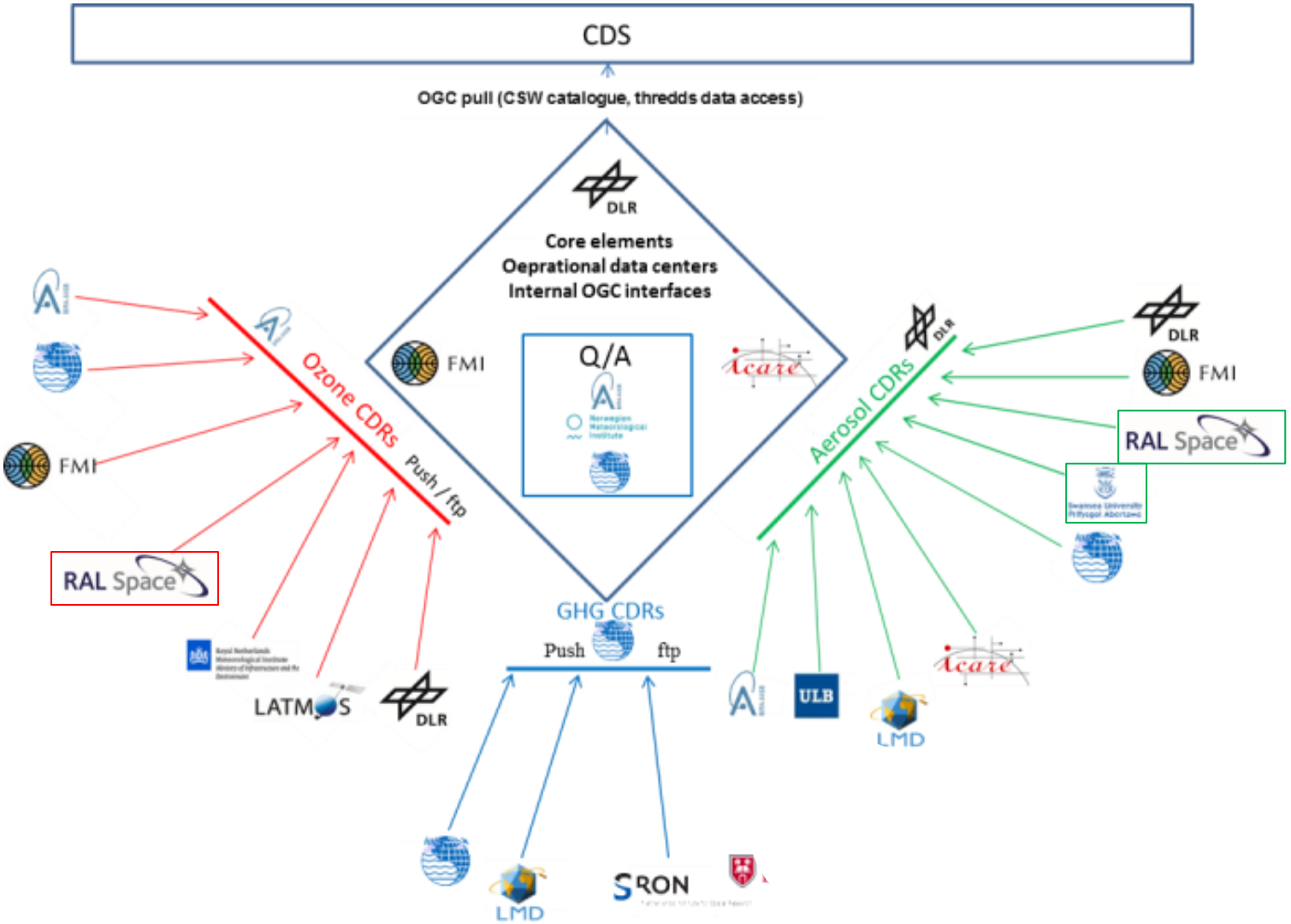
1.1 System elements and interfaces

The system of C3S2_312a_Lot2 as shown in Figure 1 is highly decentralized on two levels: First, based on scientific expertise as well as historical development and specific differences in the required input datasets and the applied retrieval algorithms, the processing is organized in three sub systems for the three ECVs (O3, AER, GHG; depicted in red / green / blue colours respectively). Secondly, within each sub system several different datasets are processed at a number of decentralised service provider institutions. These are responding to the needs for a set of variables to be provided for each ECV, exploiting measurements from different appropriate instruments and applying different suitable algorithms.
Note that providers highlighted with a box (RAL, Swansea University) have changed to in-kind contributors providing brokered datasets while the Co-PI for the respective ECV (O3: BIRA; AER: DLR) supports these data records and their documentation as custodian.

For the output of all three ECV processing systems one common interface to the Climate Data Store is implemented at DLR at http://wdc.dlr.de/C3S_312b_Lot2/. This interface is only provided for the Climate Data Store (not for users), to access all Lot2 data from it. Users will access the data through the Climate Data Store https://cds.climate.copernicus.eu/cdsapp#!/home, and from there select the specific links for the three ECVs contained in Lot2:

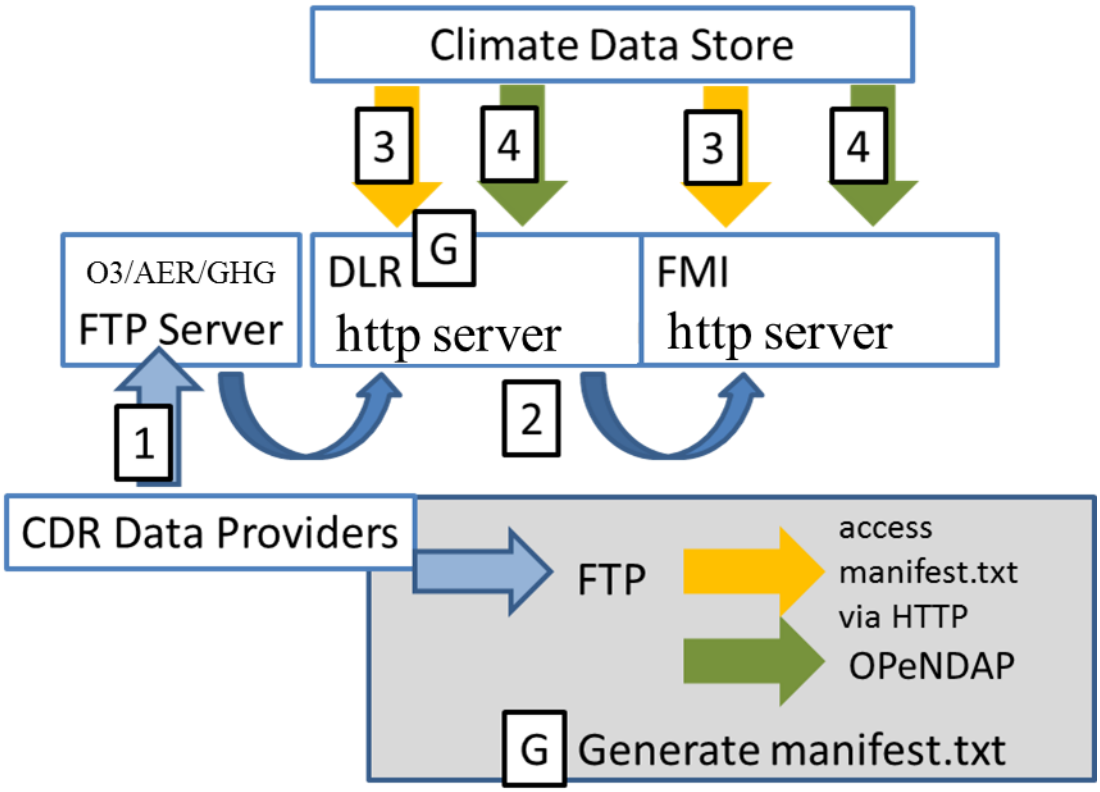
- Ozone:
  https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-ozone-v1?tab=overview
- Aerosols:
  https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-aerosol-properties?tab=overview
- Greenhouse gases:
  - Carbon Dioxide
    https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-carbon-dioxide?tab=overview
    Methane https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-methane?tab=overview

Figure 1: Decentralised structure of the C3S2_312a_Lot2 system (in-kind providers (i.e., RAL & Swansea University) highlighted with boxes).



The procedure for uploading newly processed data to the central data server is summarized in Figure 2.

Figure 2: Architecture overview of the C3S2_312a_Lot2 system.



Each individual data provider conducts their processing as scheduled and then uploads the resulting data files to the central ftp node [FTP] for the respective ECV (O3, AER, GHG) at BIRA, ICARE and Bremen university [1], where technical format control and completeness checks are conducted. From there, DLR copies the complete data collection for an ECV at a delivery point in the schedule to its http server [2], from which the Climate Data Store (CDS) can pull a dataset with its full directory path and filename [3]; the list of all files with their complete access paths is published at the same server (top level) within a computer readable *manifest.txt* file automatically generated at the HTTP server on each addition of new datasets [G]. In future, the CDS can also automatically select subsets of each data file as requested by a user with an openDAP protocol [4]. A backup copy of the complete http structure will be maintained by FMI and can then be accessed as backup also by the CDS.

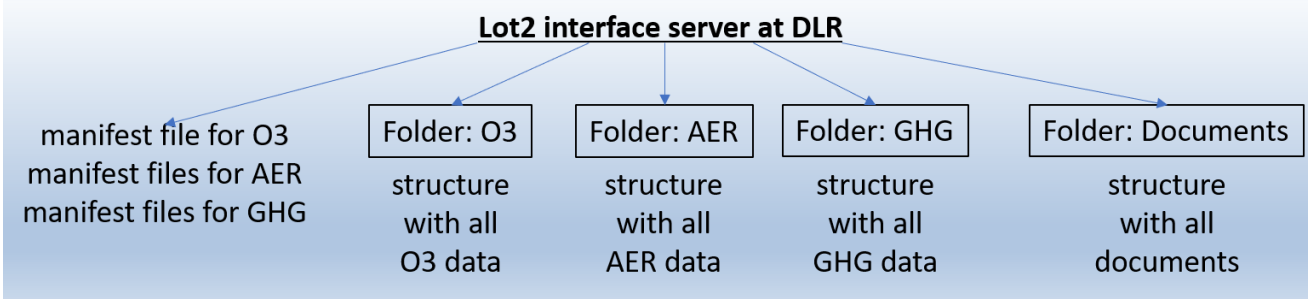
Reflecting the decentralised nature of the Lot2 contract and keeping granularity of the data holdings congruent with scientific expertise, there are separate top level folders for the three ECV collections O3, AER and GHG and also their *manifest.txt* files are kept separate.

In addition, one sub folder of the DLR server contains the latest approved versions of the documentation which accompanies the datasets.

The three specific Annex documents (one for each ECV: ozone, aerosol, greenhouse gases) to this overview document provide all information that is specific for the three ECV sub systems and for each individual processing chain of them (datasets: input, auxiliary, reference datasets; hardware with its maintenance and upgrade cycles). Here only the central adaptor / interface to the CDS is described and general principles are summarized.

Altogether, the DLR interface server has the structure as depicted in Figure 3. The datasets are contained in three specific folders for ozone (O3), aerosols (AER) and greenhouse gases (GHG). There is one additional folder which contains all documents. Those four folders have an appropriate structure of sub directories. At the level of the four folders, also the manifest files for each ECV are provided, which are the main tool for the CDS to access the Lot2 data interface.

Figure 3: Structure of the C3S2_312a_Lot2 interface server at DLR.



There are two manifest files for aerosol and for greenhouse gas datasets which group the respective products according to different naming conventions applied:

- Aerosols:
  - L3 gridded daily and monthly products (total column and layer height)
  - L3 5-daily gridded products (stratospheric extinction)
- Greenhouse gases
  - L2 products
  - L3 products

All earlier versions of the manifest files are kept and are tagged with their date.

2. Upgrade cycle implementation procedure

ECV retrieval is conducted at distributed processing nodes, operated by the partners. Each partner is fully responsible for the ECV retrieval and processing chains run at their site. This also accounts for the hardware upgrade cycle.

When negotiating the project each provider was equipped with sufficient hardware capabilities to conduct the scheduled regular extensions of existing data records (ICDR processing) and full mission reprocessing (CDR).

Each provider re-uses existing hardware from a larger satellite retrieval processing environment at their institution as in-kind contribution. This implies that no specific costing was calculated for scheduled maintenance or repair of system components, but such system upgrades need to come within the normal cycles of each institution.

On this basis with currently available systems (non-exclusive for this contract) service providers estimate their best effort capabilities and also possible peak capabilities in case of arising needs to make allowances for delays due to retrieval issues.

3. Procedures for reprocessing CDR's

In this Lot2, different procedures for the processing of ICDRs and the reprocessing of CDRs apply for the three ECVs ozone, aerosol and greenhouse gases. For ozone and aerosol, regular extensions (every 3, 6 or 12 months, depending on the different products) of existing CDRs with additional periods are achieved through ICDR processing and ultimately, a gap filling major reprocessing is applied to achieve a fully consistent new CDR. For greenhouse gases, annual reprocessing of full CDRs is conducted which includes also an extension of the periods covered.

For ozone products different datasets have different frequencies of ICDR processing (from 3-monthly, via 6-monthly to 12-monthly); the last major reprocessing, covering up to the end of 2020 was conducted in early 2021. The next gap filling / major reprocessing is scheduled for end of 2023. Where needed for consistency of algorithms or input data, complete reprocessing was done; otherwise checks for consistency and gaps were made and gaps were filled where input data are available.

For aerosol products datasets have a common frequency of ICDR processing (6-monthly); the last major reprocessing was conducted in late 2020 / early 2021, covering up to the end of 2020. The next gap filling / major reprocessing is scheduled for end of 2023. Where needed for consistency of algorithms or input data, complete reprocessing was done; otherwise checks for consistency and gaps were made and gaps were filled where input data are available.

For greenhouse gas products, the last full CDR reprocessing was conducted in 2021, covering up to mid-2020; no incremental data record extension is conducted for greenhouse gas data records, but two full CDR reprocessing activities are scheduled, at annual intervals (August 2022, August 2023). Each full reprocessing activity means a complete reprocessing of the entire records with new algorithm versions (if available) and also includes an extension of the covered period up the end of the last calendar year.

As a general principle, reprocessing is only conducted when the outcome is a new dataset (e.g. new algorithm version, new input dataset version, new output format, new gridding algorithm). In particular this contract does not include resources for algorithm improvement so that they must come from external funding and cannot be assured (e.g. CCI program, national, institutional sources). From our experience, most providers managed to secure some level of funding for this, but it has not been sufficient for major algorithm developments. On the basis of known activities, a processing schedule was defined in the contract negotiations. Before starting each re-processing, the consortium checks that processing makes sense (proper input data, some relevant change of one component). If this is not the case (either erroneous or unavailable input data or no change to the retrieval algorithm) a reprocessing is postponed and the reasons for it discussed with ECWMF.

Another reason for a (partial) reprocessing may be the result of the product validation (by the providers themselves or by the independent validation partner per ECV), which may define the need for an ad-hoc reprocessing which is conducted insofar as resources allow.

4. System maintenance and system failures

ECV retrieval is conducted at distributed processing nodes, operated by the partners. Each partner is fully responsible for the ECV retrieval and processing chains run at their site. This also accounts for the hardware upgrade cycle.

When negotiating the project each provider was equipped with sufficient hardware capabilities to conduct the scheduled regular extensions of existing data records (ICDR processing) and full mission reprocessing (CDR).

Each provider re-uses existing hardware from a larger satellite retrieval processing environment at their institution as in-kind contribution. This implies that no specific costing was calculated for scheduled maintenance or repair of system components, but such system upgrades need to come within the normal cycles of each institution.

On this basis with currently available systems (non-exclusive for this contract) service providers estimate their best effort capabilities and also possible peak capabilities in case of arising needs to make up delays due to system failures.

In case of system failure of DLR facilities providing the main access point for the CDS to all data under this contract, the server at FMI will act as backup and ECWMF will be notified accordingly. It should be noted that this server component at DLR is part of the central DLR IT infrastructure maintenance and therefore the risk of a long-term complete failure is regarded as low.

5. User support

The central user support at ECWMF for the C3S is the access point for users, and they respond to all first level technical queries. More specific queries are transferred to the responsible contracts such as Lot2.

All questions and comments related to Copernicus data products

(including the ones described in this document)

should be submitted via this website:

http://copernicus-support.ecmwf.int

Within Lot2 DLR, supported by the Co-PIs for O3 and GHG (BIRA, Bremen University) is the specific contractor that coordinates responses to user queries. For this purpose the following three dedicated user support accounts have been set up (i.e. registered with ECMWF Software Support (JIRA Service Desk)):

C3S Ozone ECV Support (c3s2_312a_lot2_ozone@dlr.de)
C3S Aerosol ECV Support (c3s2_312a_lot2_aerosol@dlr.de)
C3S Greenhouse Gas ECV (c3s2_312a_lot2_ghg@dlr.de)

The escalation of a user query from general to specialised support, by a support agent from the Copernicus User Support (CUS) Desk, occurs when a query is assigned to the appropriate Copernicus User Support (CUS) label (c3s_c3s312bLOT2_ECV-ozone, c3s_c3s312bLOT2_ECV-aerosols or c3s_c3s312bLOT2_ECV-GHG). A notification of the query escalation is automatically delivered to the appropriate mailbox (associated with the user support account linked to the appropriate CUS label).

The three mailboxes are continuously monitored, so that any level 2 support activities are conducted during the usual working hours (weekdays between 9am and 5pm, response for 85% of tickets within 5 working days).

Appendices

There are three separate Appendices to this document covering:

Appendix A: Ozone products
Appendix B: Aerosol products
Appendix C: Greenhouse gas products

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: https://support.ecmwf.int/