Copernicus workshop on climate observation requirements, ECMWF, 29 June – 2 July 2015.

Workshop summary and outcomes

There were around 20-25 participants in each of the three working groups

Working Group 1 Chairs: Mark Dowell and Carolin Richter
Rapporteurs: Gianpaolo Balsamo and Baudouin Raoult

Working Group 2 Chair: P. Lecomte
Rapporteur: H Hersbach

Working Group 3 Chair: Jörg Schulz
Rapporteur: Paul Poli

Defining the observation-based contents of the Climate Data Store (CDS) is at the heart of the discussions in this workshop. The geographic scope for the CDS is the Global and the European domains. There is a diversity of End-Users envisaged for the CDS: Policy makers, downstream application and scientific community users, etc., and these were well represented in the working groups.

The Copernicus Programme is a new and ambitious initiative of the European Commission. ECMWF is the operator of two of the Copernicus Services: the Copernicus Atmosphere Monitoring Service (CAMS) and the Copernicus Climate Change Service (C3S). The C3S is in its "proof-of-concept" phase in 2015-2016, and will develop into pre-operations in 2017. Copernicus will provide operational services and will support science indirectly. Open and free access is ensured for the data sets provided by the EU-funded Copernicus services. There will be a strong evaluation and quality control function with science support. The evaluation of uncertainties in all aspects has been emphasised.

Topic A: Collection and processing for in situ data for atmosphere, ocean, land, cryo- & bio-spheres.

For C3S reanalysis the scope of data rescue is both global and regional, with a focus on the areas of European interest (the Continent of Europe, the North Atlantic, the Mediterranean, Northern Africa and the Arctic). The data requirements vary for different climate products (reanalysis, gridded products, time-series, etc.) and depend on the societal impact areas that C3S will address.

C3S can help identify the gaps between the requirements and the actual availability of observations. The data gaps can be geographical, or pertain to a specific period of time. The C3S needs also to think about future requirements and what will become the requirement in 10-years' time. This has implications on the sustainability and enhancement of observing networks.

The links between Copernicus components was raised as integration can avoid unnecessary duplications and crosscutting activities across the Copernicus Services are perceived as crucially important.

Recommendation: ECMWF/EC to ensure integration of C3S in the wider Copernicus context, with identification of the crosscutting activities.

During the workshop discussions there was often the need to explain terminology, e.g., what is meant by 'operational', different 'levels' of data records (FCDR, etc.) or types of meta data. Copernicus should create a lexicon that clearly describes the meaning of such terms, with applicability across services.

Recommendation: C3S to create a lexicon that describes the meaning of specialist terms, with applicability across services.

A1. What activities are needed to support data rescue and collection?

There are 2 types of data rescue activity:

i. Providing easier access to data that exists and enhance its visibility, and
ii. Data mining for observations that need to be digitized.

To investigate the frequency of severe events, the data record (time series) shall enable extremes to be adequately qualified with both historical time-series and century-long reanalysis.

- C3S should create an inventory of in-situ data sets, make an impact/gap analysis and focus on rescue of the most critical data sets. Such an inventory is needed to prioritize, avoid duplication, and help make data visible. This may become an inventory of inventories. The inventory should cover rescued data and data to be rescued for all variables and provide accurate metadata (e.g., station position with an error of less than 100 m)
- C3S should provide a standard catalogue whereby anybody willing to do so can make their data available to Copernicus and thereby to the community. Standards should apply also to quality control (QC) and flagging. The standards should allow for old data of lower quality to be retained, and appropriately annotated.
- C3S should promote traceability and provide access to raw data.

- Promote the provision/use of meta data (building on ERA-CLIM activities). ERA-CLIM2 project has created an inventory of rescued datasets for upper-air and surface observations. The metadata record includes location, period, variables observed, WMO number, PI responsible. There is a similar initiative for ocean data carried out by EMODNET and the Mediterranean data portal.
- Promote the representation of uncertainties in observational data.
- Copernicus should have as one of its aims to provide a political push for free data. It is the responsibility of C3S to influence the policies in Europe towards free and open access to relevant climate data. Copernicus should use the weight of the EU to push the data policy issue forward. Clear support is provided in the recently adopted WMO resolution.
- Work with other entities (e.g. National Funding to promote activities in developing countries.

The ERA20C/20CR and ACAD experiences have shown that close connections between data recovery and reanalysis activities have advantages as the data is being quickly used and feedback rapidly provided.

Data records for ozone and aerosols are also very well established and these are of interest to reanalysis of the Earth system. Environmental and phenological data and forest inventories that can also be of interest as those are indicators of climate change that are intuitive (natural indicators). Link clearly with NCEI terrestrial efforts

Recommendation: Collect a list of key data holders (e.g. Data Centres, data ownerships) for the ECVs that will be an important resource for both meteorological and more widely environmental data. GCOS has those lists among the resources.

Recommendation: Consider producing a white paper on the added-value of sustained data rescue activity in century-long reanalysis and observation based dataset to characterize climate extremes and a strategy document proposing how activities should be undertaken in a sustainable way.

A2. What are the requirements for homogenized and harmonized in situ data records?

Data homogenization and harmonization are often research activities. Consensus and best practice could migrate to operational activity, however at sub-daily and daily scale there seems not yet to a consensus on methodology. C3S has the responsibility to influence the priorities for data rescue activities. Homogenization can be supported by the availability of re-analysis feedback information. It will be necessary to homogenize also the representation of data uncertainty – this is essential for the creation of higher-level products.

The metadata is key for ensuring consistency of time series data. In the longer term, sharing knowledge on the observing protocols may be an important step to achieve consistent and homogeneous data records.

Recommendation: Establish the concept of best practice for consistent in-situ data processing from raw-data to “climate-quality” (e.g. compliance, QC/QA, record-length) and help ensure the availability of complete and consistent metadata.

There is a need to serve data at three levels: raw, harmonized and homogenized data. The adjustments should be provided with the original data to make the process reversible.

Recommendation: Develop and propose a minimum set of metadata (e.g. including type of processing) required for quality control and effective use of the data.

A3. How can access to national holdings of in situ climate data be improved?

For data availability within Europe there might be leverage with national institution. There is WMO-leverage for data access at global scale (GAW, GCW, …). For Glaciers, contacting an International data-centre may obtain a more rapid access to the data.

Access to terrestrial in-situ data seems a more complex issue. The GTOS is currently lacking funds for coordination of terrestrial observing networks, however there are several on-going data coordination activities, and GCOS has listed the existing networks. This shall be a first point of contact. Also meteorological weather data is subject to gaps.

The data accessibility might be achieved by data exchange (under the framework of a project such as ERACLIM2). There is a need to work at multiple levels, national, institutional, and sometimes personal level to ease data sharing and data redistribution rights (e.g. EUMETNET, while being successful on data collection, has had limitations on data redistribution).

GMES had an in-situ component that is not as prominent in Copernicus but coordination of national data is embedded in all services (especially for in-situ that are Service-specific). Copernicus is a joint program with the national level activities.

Recommendation: Involve EU NMSs in data sharing activities in order to agree on the willingness of data exchange for C3S and help setting up actions.

There is a need for C3S to connect with other data and associated communities e.g. for phenological and socio-economic indicators

B: Collection and reprocessing for (Level-1) satellite data records

In satellite terminology the Level-1 covers a range of different degrees of processing, typically produced at satellite agencies. A Fundamental Climate Data Record (FCDR) is a more general category than Level-1 and can be considered beyond a single satellite agencies mandate.

For Copernicus we consider the FCDR to be the most appropriate term to refer to these satellite products. An FCDR is aiming at harmonized time series (even across different sensors and satellites) and higher requirements on uncertainty information (and even an ensemble FCDR can be requested). FCDR can have recorded within it the underlying satellites/sensors with a view to use that information in reanalysis.

B1. What are the priorities to support satellite data rescue?

There is a matter of urgency in rescuing old satellite data. Our ability to read old data may be lost. The physical location and status of archives of early data sets is not always known. The availability of specific data experts and the ability to transfer their knowledge will eventually be lost.

Landsat and AVHRR are important datasets, and there is urgency in collecting those data due to digital/magnetic tapes deterioration. For AVHRR data might be available in different institutions but there is value in a data rescue aiming at a single repository. The GRISST has collected SSTs (based on best effort).

FCDR creation is not complete: ESA CCI as a focus on Level2. Stewardship for AVHHR (low and high spatial resolution is not ensured. The generation of combined FCDRs, e.g. scatterometry reprocessing across agencies, is not coordinated. Feedback information from Level2 and reanalysis needs to be systematically used to further improve FCDRs

The collection of metadata (pre-launch characterization) might also raise some issues. Always include information on calibration and other relevant documentation. There is a need to reread original telemetry and create better meta data. Ensure that for current missions we collectively do better than for early satellite data when it comes to capturing and recording meta data.

Recommendation: Starting from Sentinel (2, 3, …) class data consider collection for the entire period of availability going back in time (e.g. Landsat and AVHRR, METEOSAT) and collecting all metadata information (e.g. including calibration and pre-launch information).

Recommendation: Encourage dialogue between satellite agencies on data accessibility using existing committee such as the CEOS/CGMS-WGClimate

Recommendation: it is recognized that long-term global data preservation is essential to support climate services. The C3S is encouraged to setup a WG that defines/ensures the priority the requirements and resources essential to deliver the services.

B2. What are the requirements for timely reprocessed product streams?

The step size of the expected quality enhancement can be a good guidance for issuing a reprocessed dataset. Typically reprocessing every year or every few years should be the target.

For SST the user community indicated yearly frequency as satisfactory. For Altimetry data the reprocessing can be less frequent. For MERIS/ENVISAT bi-annual (or annual).

The Interim Climate Data Records (CDRs) should aim at quick access. For SST it is about 2 days. The ICDR can support the reanalysis projects aiming at delivery a few-day behind the NRT operational data exchange. It is desirable to keep the ICDR processing as close as possible to that of the CDRs.

There is a need for ICDR use cases to determine where the priority for new resources should be placed. Activities where ICDRs may play an important role include heat wave attribution and 'annual state of the climate' reports. The cost/benefit of L1 ICDRs should be analysed - it will not be possible to produce ICDRs for all ECVs, and depends on user requirements.

Recommendation: Consolidate the list of variables that are in CDR to support time-critical climate applications (e.g. atmosphere, ocean and land monitoring applications) and including meteorological data/products that did not enter NRT.

B3. How should C3S link to international co-ordinating activities in this area?

For Copernicus there is a need to reach out to the other coordinating bodies and satellite agencies world-wide, for example: GSICS (for inter-calibration), SCOPE-CM, CEOS/CGMS WG-CLIMATE. OBS4MIPs-WCRP, WGClimate. Consider the added-value of linking with GEO-ring e.g. in collaboration with NOAA and JMA – maybe fulfilled by EUM.

Initiatives such as GSICS and SCOPE-CM are helpful in sharing processing tasks around the world. Such projects may or may not be effective in producing what is needed by C3S at the right time. There is therefore a need to assess if this requires European activity on non-European data.

At present there is a lack of an "international umbrella" for work on describing the uncertainties in the FCDRs. There are however two European projects looking into uncertainties (FIDUCEO, QA4EO).

There are also regional multi-national entities to engage with (e.g. Arctic, Antarctic, Africa).

B4. Are there any access issues for satellite datasets?

- The main problem seems to be the data policies. There needs to be reciprocity for the bilateral agreements.
- Access to original instrument characterisation, e.g., spectral response function can be an issue;
- Access to data from other countries, e.g., China is sometimes difficult. Third party services as from EUMETSAT can help but archiving of the data needs to be ensured for later use
- Copernicus should assess what is needed in terms of third party data and discuss with EUM/ESA ways of providing and archiving them;
- There is an increasing number of private companies proposing to sell climate data with new sensors. This has the potential to become a threat in the future and need to be considered by Copernicus to ensure access;

Topic C: Observational ECVs and gridded products

The priority in C3S is for gridded ECV products that serve sectorial applications. A useful exercise is to take available non-EU gridded products (e.g. for US datasets) and match equivalent EU products. This will involve capacity building in some areas. The added value of C3S will include improved sustainability, synthesis among products, QC and QA, etc.

Within the sectors there are community preferences on data sources based on research or services applications (e.g. use of in-situ vs satellites). EUPORIAS, EUCLEIA and SPECS projects provide examples and information on data use and sectoral requirements. It is suggested that C3S takes full benefit from user requirement collections performed by GCOS, EU FP7 and H2020 projects, CCI and SAFs. It is seen useful to compile user requirements for C3S in a systematic way.

C3S shall remain in capacity of selecting (based on compliance) and providing access (according to its mandate) to multiple datasets independently from their funding origin.

Evolution from research to operations: The Copernicus and the Horizon2020 programmes will work in synergy with a cycle of multiple years (5-years or so). Therefore an "Inner-Service evolution" component has been defined within Copernicus for developments that are associated with algorithm updates, satellites-adaptation, and short-term new developments that respond to user needs and from operational experience.

C1. What datasets are currently available and how could they be used for climate services?

There are dataset which are supported by programmes: ESA-CCI products (with R&D scope but no operational plans), EUMETSAT-SAFs products (with a CDOP phase), ECMWF-reanalyses and non-EU reanalyses.

There is a number of Framework-Program, Horizon 2020 and National R&D complementary products, such as GPCC precipitation products, GRDC runoff and river-discharge dataset and EUSTICE (H2020).

There are national and regional datasets (both R&D and Operational): the E-OBS dataset at European scale has a large number of observations gridded at 0.25x0.25 degree. The Alpine dataset on precipitation (1900 onwards using historical time series), France, Germany, Austria, Italy, Slovenia joint initiatives for data provisions. CARPATCLIM including Carpathian countries (10 km daily) general Meteorological quantities, HADOBS from United Kingdom, a glaciers data set from Switzerland and Permafrost from Germany AWI. The DWD climate dataset (www.gcos.de/inventarbericht) provide an example of National inventory with an assessment of sustainability, Spain02 gridded product from Santander University and AEMET, NCEO products, TAMSAT, Leicester University (GlobTemperature). (Noe, this list is not exhaustive). The GCOS national coordinators have up to date lists of the ECVs available nationally.

CDS shall contain precipitation products as it cuts across application sectors:

- Although ground-based radar data seem very useful to analyse extreme precipitation and storms at high spatio-temporal resolution there is no radar climatology and no exact knowledge of radar data holdings; (ongoing efforts in the US and Japan)
- No European global satellite climatology – CM SAF is starting but more support is needed, e.g., for realising ICDR;

- In situ data with increasingly higher temporal sampling exist but exchange of data is rated poor and data rescue could help;

Recommendation: Verify when GCOS national coordinators have up to date lists of the ECVs national reports, and encourage this mechanism (e.g. 3 nations Switzerland, Germany, Austria).

Recommendation: for non-EU gridded products matching equivalent EU products may identify where capacity building is needed in some areas.

Recommendation: The transition for "mature" products to C3S should also include C3S feedback to the developments for ECVs (International Agencies and National programs). This is desirable also for other Copernicus services.

Recommendation: Ensure an active dialogue with on-going H2020 projects of relevance for C3S and provide feedback.

Recommendation: The evolution and development pathway will be crucial for C3S and elements of evolution should be assured and contained within the service.

Recommendation: The consolidated list of ECVs should make sure that within the operational context there is adequate funding in C3S/Copernicus (for continuity).

C2. What kind of input data, tools and activities are needed to support further development and production of these datasets?

Infrastructures and interfaces will facilitate data access. In order to achieve that, there is need for standards in the datasets for both metadata and formats (e.g. NetCDF, GRIB, etc.).

Applications will be included in a toolbox and will be rely on standards (units, dimensions, naming conventions, date & time stamps). The toolbox will also be built on software standards. Features of the toolbox shall include: Re-gridding, format conversion, point and areal extractions, aggregation, space and time averaging and provision of statistical moments, calculation of indices etc. Traceability of uncertainty and propagation will be also part of the toolbox. Note however that there are specific examples for which there is not a unique methodology (e.g. regridding+conservation or regridding+uncertainty-mapping).

Traceability shall also allow distinguishing data sources (gridded-observational data, reanalysis).

For global data collection managed in non-EU countries the continuity of access to raw data may not be trivial in some cases (as rely on international collection efforts, e.g. ICOADS). The data access continuity may be limited to what achievable by the dataset providers (e.g. for enhance level of continuity establishing a service-level agreement with providers might be needed? Can C3S provide support international data collection?). This will need to be considered on a case-by-case basis.

The dataset will need to be supported by different level of information based on the level of expertise of the user. This is linked to metadata. Reference publications (with DOI) and the dataset itself (with associated DOI) will provide traceable information.

Recommendation: Ensure availability of DEM, Land-Sea mask, country boundaries, coastal information (and other ancillary based on best available dataset and Copernicus-Land products) for all CDS and sectoral applications.

Recommendation: Reference datasets for Calibration/Validation and Uncertainty estimates shall be part of C3S input data.

Recommendation: Provide DOIs for all datasets on the CDS. Many satellite data sets have DOIs already. All products created with C3S funding shall be provided with a DOI from C3S. Issues exist with continuations in time (ICDRs) or when metadata change: C3S should engage in the definition of a consistent approach.

C3. What could be the role of Copernicus (and C3S in particular) in facilitating this development?

Recommendation: Traceability (at large) shall be part of the CDS provided information (also in reanalysis via EO-feedback)

Recommendation: For international data collection of strategic interest to C3S/Copernicus has to consider continuity aspects and support (e.g. ICOADS-type)

Recommendation: Organize/support workshops to identify future needs for users and research, identify best practices, and advance C3S capabilities, e.g.

- Bringing together researchers and users
- Visualization communities
- Homogenization communities
- Within the EQC

D: General issues

D1. What quality/maturity criteria should be applied to candidate datasets for the Climate Data Store?

System maturity matrix (Core Climax) already defines suitable quality criteria. The quality assessment should not be done by the data producers but by the C3S. The CHARMe project provides good examples. Data sets should be associated by lists of publications, associated website, DOI and a short guidance note (~5 pages). C3S should ask providers to produce such sort notes.

For inclusion in the CDS, there needs to be criteria for selection and a long-term commitment for maintenance and production of the data sets. It is suggested that entry into CDS may be granted for data sets achieving maturity levels 2-3 in all categories. However, for each variable needed in the CDS there should be at least one data set independent of maturity assessment result (e.g. historical data, even if below maturity level 2).

The idea of a maturity index may not be applicable for non-satellite datasets. The level of maturity/quality is ECV dependent. Data maturity matrix must be complemented with an application maturity (fit for purpose) matrix. ECV products should have realistic uncertainty information attached to them, to enable the user to determine whether or not the data are of sufficient quality for their need.

Document existing quality/maturity measure/criteria and communicate them to the users. Make sure gaps are explained.

Recommendation Document existing quality/maturity measure/criteria and communicate them to the users. Make sure gaps are explained. Make sure that all datasets have similar level of documentation.

Recommendation: Reference dataset are requirement for C3S CDS, with more strict quality requirements.

Recommendation: Data maturity matrix must be complemented with an application maturity (“fit-for-purpose”?) matrix.

Recommendation: ECV products should have realistic uncertainty information attached to them, to enable the user to determine whether or not the data are of sufficient quality for their need.

D2. What is needed to achieve open access?

It was not clear what is meant by “open access”. According to its definition on https://en.wikipedia.org/wiki/Open_access

- *Gratis*: free and unrestricted
- *Libre*: free with attached license (e.g. creative commons)

Licenses often impose conditions (no-redistribution, cite owner,…) This can be seen as a hindrance. Is registration against open access?

Current plan for the CDS sees no login required until downloading data or running process. The workshop recommended C3S to work with national agencies and other data providers towards establishing open data policy for data served via the CDS.

D3. What are the requirements for metadata (relevant to WIGOS)?

The requirement for discovery metadata is to follow ISO-19115 (INSPIRE, WIS, GEOSS et cetera). Metadata records must be in UTF-8. Small differences between meta data requirements in WIGOS, INSPIRE, CF and also for DOI registration may need to be ironed out. Meta data standards shall be the same throughout all Copernicus services;

The WG agreed that quality criteria shall also apply to metadata (e.g. completeness). Meta data definitions need clarification across sectors, users e.g. Commentary metadata (User feedback). Note that the recently upgraded OSCAR (WMO) provides information on global observational data sets distributed under the auspices of the WMO. Uncertainty is regarded to be part of data set, so not as metadata.

D4. What are the observation requirements for evaluation of climate models? What to compare with the simulations, for past, present and future?

Observations are needed for the monitoring of model performance and for the continuing improvement and development of models, i.e. for the evaluation of the model representation of physical processes.

There are existing lists in research communities. The C3S requirements should be aligned with those of CMIP. The WG commented that there might be observations beyond ECV as in principles other measurements in the earth system can be of interest (therefore the list is non-static). There is a need for uncertainty information/estimates associated to validation products.

Tentative list of ECVs/indicators to be covered by the C3S

Surface air temperature	*Ocean colour*	Snow cover
Surface precipitation	*Sea ice*	*Glaciers & ice caps*
Water vapour	*Sea level*	Albedo
Surface radiation budget	*Sea surface temperature*	FAPAR
Earth radiation budget	Global ocean heat content	*Fire*
Carbon dioxide & methane		*Ice sheets*
Ozone & aerosols		Lakes
Cloud properties	CO_2 partial pressure	Permafrost
Wind speed & direction	Ocean acidity	*Land cover*
Upper air temperature	Sea surface salinity	Leaf area index
Other long-lived greenhouse gases	Current salinity	*Soil moisture*

ECVs & indicators (typo Current salinity → Ocean current & Ocean Salinity)

Surface air temperature	*Ocean colour*	Snow cover
Surface precipitation	*Sea ice*	*Glaciers & ice caps*
Water vapour	*Sea level*	Albedo
Surface radiation budget	*Sea surface temperature*	FAPAR
Earth radiation budget	Global ocean heat content	*Fire*
Carbon dioxide & methane		*Ice sheets*
Ozone & aerosols		Lakes
Cloud properties	CO_2 partial pressure	Permafrost
Wind speed & direction	Ocean acidity	*Land cover*
Upper air temperature	Sea surface salinity	Leaf area index
Other long-lived greenhouse gases	Current salinity	*Soil moisture*

The GCOS Essential Climate Variables (extracted from GCOS portal)

There are 50 GCOS Essential Climate Variables (ECVs) (in the 2010 updated report) required to support the work of the UNFCCC and the IPCC. All ECVs are technically and economically feasible for systematic observation. It is these variables for which international exchange is required for both current and historical observations. Additional variables required for research purposes are not included in this table. It is emphasized that the ordering within the table is simply for convenience and is not an indicator of relative priority.

The GCOS Essential Climate Variables are divided by sector:

Atmospheric (over land, sea and ice)

- Surface:[1] Air temperature, Wind speed and direction, Water vapor, Pressure, Precipitation, Surface radiation budget.
- Upper-air:[2] Temperature, Wind speed and direction, Water vapor, Cloud properties, Earth radiation budget (including solar irradiance).
- Composition: Carbon dioxide, Methane, and other long-lived greenhouse gases [3], Ozone and Aerosol, supported by their precursors[4].

Oceanic

- Surface:[5] Sea-surface temperature, Sea-surface salinity, Sea level, Sea state, Sea ice, Surface current, Ocean color, Carbon dioxide partial pressure, Ocean acidity, Phytoplankton.
- Sub-surface: Temperature, Salinity, Current, Nutrients, Carbon dioxide partial pressure, Ocean acidity, Oxygen, Tracers.

Terrestrial

- River discharge, Water use, Groundwater, Lakes, Snow cover, Glaciers and ice caps, Ice sheets, Permafrost, Albedo, Land cover (including vegetation type), Fraction of absorbed photo-synthetically active radiation (FAPAR), Leaf area index (LAI), Above-ground biomass, Soil carbon, Fire disturbance, Soil moisture.

---

[1] Including measurements at standardized, but globally varying heights in close proximity to the surface.

[2] Up to the stratopause.

[3] Including nitrous oxide (N2O), chlorofluorocarbons (CFCs), hydrochlorofluorocarbons (HCFCs), hydrofluorocarbons (HFCs), sulphur hexafluoride (SF6), and perfluorocarbons (PFCs).

[4] In particular nitrogen dioxide (NO2), sulphur dioxide (SO2), formaldehyde (HCHO) and carbon monoxide (CO).

[5] Including measurements within the surface mixed layer, usually within the upper 15m.