

## WCRP Task Team on Seamless Data and Data Management

### Input to the Implementation Plan - Report to JSC

2 December 2019

Climate observations and climate models are producing significant amounts of data and information and represent an intrinsic and critical part of WCRP activities. The purpose of this Task Team was to develop **recommendations on mechanisms and structures** needed so that the WCRP can achieve **integrated data production and management** across the Programme.

We understand seamless climate data to be all data required for understanding, predicting and projecting the climate system across all Earth-System components and scales. We understand seamless data management to be the coordination of continuous and interconnected production, acquisition, processing, archiving, transmission and dissemination of climate data across all disciplines, activities and scales.

#### 1) Current status of Seamless Data and Data Management within WCRP

**Data panels within WCRP core projects** (e.g., GDAP in GEWEX, GSOP in CLIVAR) carrying out fundamental work for developing, assessing and updating a number of reference data sets

**Data infrastructures** being developed as part of several initiatives (S2S, CHFP, ESGF, etc), but not necessarily interfaced or using seamless formats and best practices

**WCRP Data Advisory Council (WDAC)** as single entry point for all WCRP data, information, and observation activities

- **WDAC: Coordination across WCRP**
  - Promotion of open data policies, protocols and standards
  - Recommend best practices for ECV (Essential Climate Variable) data set development and assessments
  - Coordination of reanalysis inter-comparison efforts and review of flux efforts
  - Promotion of observational and reanalysis data to support climate modelling (e.g., obs4MIPS, ana4MIPs, CREATE-IP)
  - Regular briefings on OSSE, data assimilation
- **WDAC: Coordination with main partners (GCOS, GOOS, WWRP, GAW, Future Earth)**
  - Linkages to GCOS: currently mainly via the joint AOPC, TOPC and OOPC panels
  - Link to satellite agencies via WG Climate

#### 2) Lessons learned and open questions

Within WCRP there is a clear **need for an integrated management** of

- Observations for process understanding: beyond typical availability of 'operational' systems (dedicated field experiments)
- Observational climate data records
- Reanalyses and data generated by climate models (need to define 'data' in general)
- Data assimilation

- Data availability via open data infrastructures (e.g. ESGF, World Data Centres): need gap analysis
- Strategy on capturing observational uncertainties/covariances
- Synthesis on data stability and quality control (need for guidelines within WCRP)
- Data science and data mining/machine learning (need for information and knowledge exchange)

Part of this is covered by WDAC (e.g., reanalyses, obs4MIPS etc.), but some aspects need a stronger focus (e.g., observations for process understanding, uncertainties and quality control).

#### **Topics missing and open questions**

- New sensors and data products (e.g., micro-satellites, citizen science)
- Research-operations synergies (data management infrastructures, observational campaign vs operational networks) – e.g. WMO GDPFS
- Training and education
- Where should fluxes go (which cover all observations, modelling and data assimilation and are a key element of the Earth system approach)?

#### **Potential action items**

- Create room for stronger synergies between data assimilation, (re)analysis work and WIGOS
- Data infrastructure such as ESGF are not fully exploited across the programme. Interfacing with other systems, including operational ones (WIS/WIGOS, CDS/C3S) should be pursued.
- Reanalyses are currently largely uncoupled and need a stronger Earth system approach.
- Create guidelines on data quality control (avoid overlap with GCOS guidelines for ECVs). Advocate publication of quality control standards which would also provide recognition for effort spent, e.g. in Earth System Science Data.
- Encourage efforts for data synthesis, data integration and quality control, including for the provision of long-term, accurate (= stable/comparable) time series of climate relevant data.
- Check strength of the archival of both contemporary and historic research data. Including 'stability' of data centres / data storage sites.

### **3) Recommendations for data activities, mechanism and structures in WCRP (including resource availability/requirements)**

- Need to coordinate observations, reanalyses, data science and data management issues across the programme and across WMO (with WWRP and GAW in particular)
- Information on (and access to) datasets via inventory for all WCRP key research? Important step towards seamless approach. Can provide direct input for gap analyses. However, this would need to be adequately resourced in terms of staff time.
- Better transfer of knowledge/experiences of/in data management across WCRP entities
- Establish a strong link to space agency bodies to exchange WCRP needs and space agency plans (involve GCOS and others to communicate requirements to space agencies)

- Data management strategies should include observations, reanalyses and model simulations seamlessly (close collaboration with modelling group)
- Strengthen coordination of reanalyses, in particular around Earth system reanalysis (TIRA white paper)
- Promote a broader Earth System approach to observations with GCOS
- Include data assimilation (OSEs/OSSEs in coordination with WWRP/DAOS/PDEF and WGNE)
- Include data science and data mining as we face huge and steadily growing amount of data (connect with AI/IT communities more closely)
- Interfacing/integrating (research) data infrastructure with their operational equivalent (WIS, C3S/CDS) is a necessary condition to the R-O goal

**The future of seamless data and data management within WCRP** requires a coordinating body. This could be a combination of 1) WDAC (with strong involvement of the existing/new core projects) and 2) a broader data group across WMO. Moving the role to the JSC or to a project office is not advisable. A number of current WDAC activities could fall under a broader umbrella within the Research Board (e.g., coupled data assimilation, Earth System reanalysis) and form a similar group to WGNE.

#### **4) Current/future linkages with GCOS/GOOS, WWRP, GAW and Future Earth**

- Interfacing between WCRP/WDAC and GCOS is crucial and should be further promoted (e.g. currently via joint panels)
- Champion 'sustained observations' for ECVs (e.g., only 20-25% of ocean observing system is sustained): WCRP needs to play a stronger advocacy role for observations
- Delineation of WDAC/WCRP vs GCOS responsibility should be clarified to avoid possible overlap and mission creep
- Rolling Reviews of Requirement for observations specific to climate science application area are hard to consolidate given the wealth of research needs and activities
- NB: The future of GCOS is still to be defined/confirmed as part of the WMO Reform
- Need to ensure that space-based observations will fit in the new WMO infrastructure
- Some linkages to Future Earth (e.g. SOLAS, AIMES, IGAC) in various forms
- Broader ambition with WWRP and GAW on data matters should be established