Summary report

1.1 Remarks from SG representative and Co-chairs
1.2 Introduction of participants
1.3 Meeting logistics
1.4 Adoption of the agenda

2. Setting the stage for Global Data Management for Climate

2.1 WMO IPET-CDMP: Background, ToR, vision and expected deliverable (IPET co-chairs),

Current Terms of Reference.

1. Develop consistent guidance around climate data management from various sources, and metadata, including on standards for preservation, archiving, management and information services; that would boost the availability, discovery and the exchange of climate data with the required quality and timeliness;

2. Undertake tasks to review and update the applicable WMO Technical Regulations and guidance documents, defining as deemed necessary new technical regulations relevant to data and metadata in support of climate;

3. Recommend training and capacity development strategies and identify the elements of the required global infrastructure;

4. Further develop the proposed redefinition of the Climatological Standard Normals as a Technical Regulation to Cg-17, and develop a communication strategy to explain the use of the new standard to Members and their key stakeholders;

5. Investigate ways to combine satellite information with station observations for vulnerable and data sparse areas by filling spatial and temporal gaps in existing climate observations;

6. Engage with other technical commissions and other groups such as WIGOS, WIS, GCOS, JCOMM in matters on climate data.

Specific areas to address:
- Derive or formulate standard definitions of basic climate terms and operations, such as defining daily minimum/maximum, completeness of time-series data for forming monthly averages, etc
- A manual of, and guidance material for, basic climate DM operations such as quality control, archival, backup and technology migration, is complete and current. This may mean, for instance, instituting a work program to complete the update of “Guidelines for the Quality Control of Surface Climate Data”,
- Developing and promulgating of a commonly agreed set of standards and consistent practices for key data management
- Identifying an extended range of climate and related data types needed to support the GFCS (the 2011 version of the Statement of Guidance to Et-EGOS will help in that regard).
- The updated Guidance material on Data Rescue that will form part of the I-DARE portal. Similarly, there may be scope for other best-practice guidelines, such as the guidelines paper I’m drafting on best-practice implementation of CDMS in developing countries.
- Identify which climate data and data management operations require the formality of a WMO Technical Regulation and draft accordingly. For instance, Climate Normals.
- Internationally-agreed data exchange products and reports, e.g., World Weather Records, CLIMAT, climate indices, Daily CLIMAT etc.
- The importance of provenance metadata to underpin time-series analysis, homogeneity etc.
- Work on combining data from different sources, e.g., ToR (5) above.
- Promote the adoption of the standards defined in the CDMS Specifications document, noting that a unified technological solution is not needed, but a commonly agreed set of standards for fundamental climate data operations.
- Linking CDMS and other climate-related infrastructure (e.g. I-DARE) into WIS.
- Collaborate with members of the Rapporteurs group on climate observations to ensure climate data needs (including those for
marine, terrestrial, cryospheric and space-based data) are accommodated within the HQ-GDMSC.

- Facilitate training and other capacity-building activities in relation to climate data management.
- defining clearly the roles and responsibilities for climate data management at national, regional and global scales.

**Coordinating activities with broader WIGOS initiative:**

- Collaborate closely with relevant aspects of the WIGOS Data Management Framework initiative. Examples might include, for instance, developing best-practice guidance on data (and metadata) management and data policy, including standards and consistent practices, establishing a common vocabulary (terminology) around data management and its elements, and harvesting the value in non-NMHS (so-called Third Party data).  
- A consistent approach to the formatting, storage and preservation of climate-relevant remote sensed and model data, and integration of this with broader data management initiatives, specialist data centres, etc.
- Similarly, integrating CDMS with more general data management operations within NMHS. This might include, for instance, strategies on backup and data security, data and information exchange protocols, system interoperability, Big Data, harnessing the power of Opensource, the Cloud, etc.

**IPET-CDMP meeting – possible working arrangements:**

- Could ask the group to identify where they see the current deficiencies in climate data management are – perhaps a brainstorming session led by a few points such as the above.
- Need to get the team to suggest ways of addressing all the issue above, perhaps establishing small task teams to progress the work.

---

1 Climate services are very varied, but not all require the highest standard of measurement, metadata, etc. This links in with the concept of tiered data, reference climate stations etc.
2.2 WMO Climate Data Management Systems (CDMS) William Wright

2.3 Congress decisions / Resolutions relevant to Climate Data

Omar Baddour informed on the most relevant Congress17 decisions, namely on Climate Data Management System, Data Rescue, High Quality Global Data Management Framework for Climate, Centennial Stations, WMO policy to support GFCS and Climatological Standard Normals. Most prominent decisions was the Congress-17 Resolution 34 on Definition of Standards for Climate Data Management Systems, which called for further development of CDMS specifications in view of developing Standards for climate data management and related systems, and that the standards be referenced in WIS regulations, i.e. Manual on WIS and the Guide to WIS and the request to develop a Concept of Operations for the HQ-GDMFC and request for guidance for securing data homogeneity when transitioning from conventional to automatic stations. Congress17 adopted Resolution 60 on WMO policy for the international exchange of climate data and products to support the implementation of the GFCS. It includes an additional list of required data for exchange in addition to those listed in Resolution 40 and Res.25 of Congress13. Part of additional data, include historical climate time-series from the Regional Basic Networks (RBCNs), the GCOS Upper-air Network, GCOS Surface Network at a temporal and spatial resolution necessary to resolve the statistics of climate, including trends and extremes. Other additional data coers ECV for the Ocean, Climate relevant coastal interface data, in particular, sea level, waves and storm surges; Data on the composition of the atmosphere including aerosols; Climate relevant satellite data and products and climate relevant cryosphere data, in particular snow cover, snow depth, glacial monitoring, permafrost and lake and river ice

2.4 The concept of high Quality Global Data Management System for Climate (HQ-GDMSC). Christina Lief made a presentation of the paper she wrote with Chenghu. The Expert Team – Climate Data Modernization Program (IPET-CDMP) is designed to strengthen the management of climate data through the creation of a HQ-GDMFC, and publication of a Manual on Climate Data Management. The overall aim of this initiative is to examine and make recommendations to improve existing practices in managing climate data, as well as making more effective use of other data forms to improve climate services, including remote sensing and climate model data. It will seek to address inconsistencies and gaps in the definitions and procedures relevant to climate data management, and provide a stronger regulatory basis for climate data management processes. The HQ-GDMFC Manual will also provide best-practice guidance on methods,
procedures, and techniques in climate data management, and support development of tools and infrastructure in relation to climate-focused activities such as data rescue and preservation; archival; procedures for quality control and assurance; calculation of statistics and climate parameters; homogenization and adjustments, among others. This initiative will collaborate closely with a wider cross-programme initiative aimed at modernizing all WMO programmes’ data management infrastructure and processes and will include quality management systems for climate data management such as change management.

2.5 Data Storage, retention and climate data infrastructure

- Australia         William Wright
- China,            Guofu Wang
- France,           Matteo Dell’Acqua
- Germany,         Lydia Gates
- Morocco,         Meriem Alaourri
- USA-NOAA-NCEI,     Christina Lief

Archival/Retention policies

1. **Australia Bureau of Meteorology**

Dr Wright presented an overview of data archival and retention issues at the Australian Bureau of Meteorology, noting that exponential increases in data from the new generation of high resolution satellites and numerical weather prediction was placing additional strain on data management systems already compromised by legacy problems such as uninteroperable and undiscoverable databases, a mindset of keeping everything forever, and a lack of understanding about roles and responsibilities for data management. He requested other speakers to identify whether they faced similar problems and how they were addressing them. He described the Bureau’s proposed response to the problem, noting there was no “silver bullet” answer, and that a mix of strategies was required. He described in detail a proposed approach to working with domain subject matter experts to help them define their data retention requirements, based on such considerations as reproducibility of the data, variations in value during the data life-cycle, likely future demand, and others. He concluded by describing the Bureau’s proposed next steps, which included surveys to determine who was the custodian of what, exploring the scope for cloud hosting and partnership arrangements to help manage data collaboratively, and requiring future data-producing activities and projects to complete data management plans.

2. **France Meteo-France**

M. Dell’Acqua presented an overview of data archival and retention issues at Meteo France. He confirmed that Meteo France too was faced with the exponential increases in
data from the new generation of high resolution numerical weather prediction and satellites and radars observations. He noted that the daily amount of data added to the archive grows exponentially in that in such a system “old” data represents only a small part of the archived volume and that the bulk of data is “recent” data, therefore deleting old data in an exponentially growing archive is somehow meaningless. He added that more than 60% of Meteo France archive is dedicated to numerical weather prediction and climate data. He described the service-oriented architecture set-up by Meteo France in order to cater for the increase of data and the new user requirements and explained the principle of Meteo France retention strategy based on value of the data for the business units, and reproducibility of the data. He concluded by explaining Meteo France’s plans in the area of data processing and management chain and in the provision of high level services on the data.

### 3. DMN Morocco

M. ALAOURI presented an overview of data archival and retention issues at Moroccan Meteorological Service (DMN). She described the current observation network and demonstrated then the large increase in the amount of climate data and the diversity of data sources by installing an interesting number of automatic weather stations, as well as the acquisition of new radars, lightning sensors, air quality stations, etc. Especially when considering the huge amount of numerical models data and Climatological archives in paper format that the DMN is committed to digitize under the data rescue program. The main problem raised is that these various climate data are managed by different entities and stored in uninteroperable and indiscoverable databases. She also described the structure of the used climate data management system whose main drawback is the database model. To deal with the significant increase in the amount of data, Moroccan Meteorological Service plans to increase its storage capacity through the acquisition of a new archiving robot and a more efficient new CDMS. She concluded that nevertheless, there is always the issue of interoperability of different systems and databases besides the lack of a strategy and a clear vision about climate data archiving and retention.

### 4. NOAA-NCEI

Christina Lief gave an overview of the National Centers for Environmental Information (NCEI)-NC (Asheville) data archival and retention policies for Weather and Climate data. After a review period of 3.5 years, a new data retention schedule was released on August 6, 2015 for the National Centers for Environmental Information (NCEI)-NC (Asheville). After consolidation in 2015, the NOAA National Climatic Data Center (NCDC) was renamed the National Centers for Environmental Information (NCEI)-NC. NCEI-NC was established in 1951 as the designated archives for weather and climate
records and data held by NOAA. NCEI-NC holdings consist of two primary types of records. Observational records are original measurements of atmospheric conditions. These records come from land-based weather stations, buoys, ships, radar installations and weather satellites. Homogenized and derived products are assembled or created from original observations to make weather information more usable and understandable. NCEI-NC also archives weather and climate model data as well as publications. All items in the schedule below are media neutral unless otherwise stated. The full data retention policy document for NCEI-NC can be accessed and downloaded at [http://www.archives.gov/records-mgmt/rcs/schedules/departments/department-of-commerce/rg-0370/daa-0370-2015-0001_sf115.pdf](http://www.archives.gov/records-mgmt/rcs/schedules/departments/department-of-commerce/rg-0370/daa-0370-2015-0001_sf115.pdf)

A summary of the NCEI-NC records schedule items for Weather and Climate data are (the vast majority of new digital data archive requests fall into the categories in **bold**):

1) **Routine Atmospheric Observations: Permanent**
2) Non-routine atmospheric observations (occasional, short-duration or research): 10 years
3) Autographic Charts (self-recording): 30 years
4) **Homogenized or Derived Products: 10 years**
5) Summaries/Reports/Bulletins
   - Routine events: 20 years
   - Historically significant events: Permanent
6) Maps and Charts
   - Plotted: 5 years
   - Analyzed: Permanent
7) Model Data
   a) Data Inputs (Data input into a computer): 15 years
   b) Model runs (Data generated by the computer representing future conditions): 5 years
   c) Analysis Data (More concise, digestible data used to analyze the future conditions predicted): 15 years
   d) Re-analysis Data (Data input back into the computer for updated weather and climate predictions): 7 years after a new analysis is archived or 15 years, whichever is sooner?
8) Climate Normals
   - Core Normals: Permanent
   - Supplemental Normals: 20 years
9) National Weather Service Products; National Digital Forecast Database: 10 years

2.6 Open discussion on data storage and retention policy from Commissions and programme perspectives

3. Current and future Infrastructure in support of Climate Data Management. Regional and Global scales
3.1 Commissions and Programmes perspectives

CCI.

Dr William Wright presented the main issues around data management from a CCI perspective, and described how OPACE 1 had established Expert Teams and Rapporteur groups to address the issues identified. The work of two Task Teams from OPACE 2, on Homogenisation and on the use of satellite data for climate monitoring, were also considered relevant to the IPET's work. The main issues around climate data were that:

- it was insecurely managed in many countries;
- it was often at risk of loss, and inaccessible;
- data management practices were poorly defined and largely not subject to binding regulations, with a lack of consistency around data management-related practices;
- climate needs for data and its management were not well defined;
- the climate program in general was not well placed to take advantage of technological advances (e.g., Cloud hosting of data and tools; Opensource development communities; Big Data management); and
- data managers were often largely ignorant of their responsibilities, and there was a need for a major capacity-building exercise to develop the relevant capabilities.

Dr Wright then summarised the essential points of difference that make the management of climate data in many ways more challenging that for data from other domains, but finished by describing the opportunities afforded by a collaborative, cross-domain approach to data management. Peer Hechler then briefly described aspects of CCI's current data management-related infrastructure, including RCCs and data rescue portals.

JCOMM.

Lydia Gates described issues and developments from a JCOMM perspective. Main points included:

- The JCOMM Expert Team on Marine Climatology (ETMC) is leading the development of the JCOMM Marine Climate Data System (MCDS), which is a modernization of the Marine Climatological Summaries Scheme (MCSS), and is intended to be JCOMM's contribution to the GFCS, as well as to eventually contribute to the High Quality Global Data Management Framework for Climate lead by the CCI. MCDS will include management of data flows from various sources through Data Acquisition Centres (DAC), Global Data Assembly Centres (GDAC), and a small number of Centres for Marine Meteorological and Oceanographic Climate Data (CMOC), providing overall data integration at the variable and end user products level.
- The International Oceanographic Data Exchange (IODE) committee of the Intergovernmental Oceanographic Commission of UNESCO (IOC) and JCOMM have
jointly established an Expert Team on Data Management Practices (ETDMP), focussing on modernising data management operations, best practices and standards.

- Existing JCOMM datasets (e.g., moored buoy, glider data) were typically non-standard, not necessarily well documented or understood or linked. There was a need to develop standard data management systems that linked all sources. Other sources provide for some level of standardization, and have good recognition and trust in the international user community (e.g. ICOADS, Argo), but lack formal recognition in the WMO.

- Higher-level Automated Quality Control of the data will be carried out using software developed by DWD, with more stringent QC carried out at the GDACs.

- CMOCs cover a relatively large scope, may have a regional or programmatic focus, and carry out data and metadata rescue, data integration by variable (DACs and GDACs being the main sources), and higher level quality control, including bias correction, and the production and delivery of integrated marine meteorological and oceanographic climate products to end users. There is a current focus on the Asia-Pacific Region with one CMOC established by Cg-17 in Tianjin, China. Plans are to establish a small number of mirrored CMOCs (less than ten).

- All data and products developed within the MCDS will be discoverable through the WMO Information System (WIS) or IODE Ocean Data Portal (ODP).

**GCOS.**

Tim Oakley presented on behalf of GCOS, noting that GCOS was developing a new implementation plan that would (should?) consider the life cycle of observations, data transmission, and data management aspects, including data rescue. It was recognised that archival and retention strategies should be developed at the system planning stage. Obtaining more data, including data at higher temporal resolution (via daily CLIMAT) will be a focus, but that if SYNOPs were to be used to derive any climate messages (e.g. CLIMAT), much better quality control would be required. Tim encouraged CCI’s proposed further development of the CDMS specifications to be informed by GCOS data management needs.

**GDPFS.**

Abdoulay Harou described the planned evolution of the Global Data Process and Forecasting System, approved by Cg XVII, noting that the System would provide prediction capability on all time-scales from nowcasting through to long-range climate prediction, in support of DRR, GFCS, WIGOS/WIS, aviation and others. A proposed long-term goal was to provide a national severe weather warning capability that all WMO member countries could draw on. In terms of data management, the main issues raised were:
- Societal impacts data/information needs to be brought into the GDPFS environment to ensure better-targeted services;
- The requirement for multi-disciplinary, multi-time-scale forecasts required an integrated approach to data management and access;
- There was a need to establish suitable retention periods for the data, including non-conventional data;
- The GDPFS looked forward to integrating its plans with the IPET-CDMP

**WIS/CBS.**

Matteo del’Acqua reminded the meeting that WIS’ mandate was about enabling Members to find/access exchange and manage data, and briefly described the relative roles of National Centres, DCPC and GISCs. He noted that if a DCPC had a data product to offer, it should be registered with the relevant GISC centre, and the associated discovery/access/retrieval metadata provided to the GISC, to enable potential users to find and access the data or product. WIS in future wishes to increase the amount of data available, which would require that distributed entities would need to be robust enough to handle the increased flow, and adequate Wide Area Network connectivity. Matteo and Steve Foreman described several of

3.2 Review of the HQ-GDMSC Concept document to be converted into concept of operation document as requested by congress

The meeting split into two groups to review the draft Concept document that is requested by Congress. Consolidation was done and a new version of the draft was produced by the writing team composed of William Wright, Christina Lief and Chenghu Sun. Omar Baddour provided the secretariat support. Due to lack of time, the final draft needed additional information and editorial work.