|  |  |
| --- | --- |
| **World Meteorological Organization**  **COMMISSION FOR BASIC SYSTEMS**  **WMO Workshop on Information Management**  Geneva, 2-4 October 2017 | **WWIM / Final report** |
| 23.JAN.2018 |

# WMO Workshop on Information Management 2-4 October 2017 Final ReporT DRAFT 3



# DISCLAIMER

**Regulation 43**

Recommendations of working groups shall have no status within the Organization until they have been approved by the responsible constituent body. In the case of joint working groups the recommendations must be concurred with by the presidents of the constituent bodies concerned before being submitted to the designated constituent body.

**Regulation 44**

In the case of a recommendation made by a working group between sessions of the responsible constituent body, either in a session of a working group or by correspondence, the president of the body may, as an exceptional measure, approve the recommendation on behalf of the constituent body when the matter is, in his opinion, urgent, and does not appear to imply new obligations for Members. He may then submit this recommendation for adoption by the Executive Council or to the President of the Organization for action in accordance with Regulation 9(5).

© World Meteorological Organization, 2018

The right of publication in print, electronic and any other form and in any language is reserved by WMO. Short extracts from WMO publications may be reproduced without authorization, provided that the complete source is clearly indicated. Editorial correspondence and requests to publish, reproduce or translate this publication in part or in whole should be addressed to:

Chairperson, Publications Board

World Meteorological Organization (WMO)

7 bis, avenue de la Paix Tel.: +41 (0)22 730 84 03

P.O. Box No. 2300 Fax: +41 (0)22 730 80 40

CH-1211 Geneva 2, Switzerland E-mail: [Publications@wmo.int](mailto:Publications@wmo.int)

NOTE:  
The designations employed in WMO publications and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of WMO concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or products does not imply that they are endorsed or recommended by WMO in preference to others of a similar nature which are not mentioned or advertised.

This document (or report) is not an official publication of WMO and has not been subjected to its standard editorial procedures. The views expressed herein do not necessarily reflect those of WMO or its Members.

# WMO WORKSHOP ON INFORMATION MANAGEMENT

# 2-4 OCTOBER 2017

# FINAL REPORT

## Executive Summary

1. The WMO Workshop on Information Management was held in Geneva from 2 to 4 October 2017. It was opened by the Chair, M. Jean, President of the Commission for Basic Systems. W. Zhang, Assistant Secretary-General of WMO, welcomed the participants and emphasized the increasing importance of information management practices for members of WMO.
2. The aims of the workshop were:

* To identify key climate-related data sources that should be promoted as reference sources and outline a project to enhance their visibility in the WMO Information System;
* To simplify user access to them and to provide guidance on the development of a WMO-wide guide on information management through a review of existing practices and guidance on information management used by WMO Programmes and other organizations.

1. Thirty-two talks were given by 30 speakers to inform breakout sessions that developed priorities and information to guide the work of the Commission for Climatology’s Inter-Programme Expert Team on Climate Data Modernization Programme and the Commission for Basic Systems’ Task Team on Information Management.
2. Key issues identified by the workshop included:

* Users of climate information needed assistance in identifying the data most relevant to their needs. The Commission for Climatology would investigate an objective approach to recognizing a limited number of climate related datasets as “best trusted” datasets as an interim measure to provide guidance while a maturity-based approach for assessing datasets was being developed and implemented.
* Areas for which guidance on information management were identified and would be used by the Commission for Basic Systems to prioritize development of standards and best practice to be included in the documentation of the WMO Information System.

1. The workshop recommended that CCl should produce project plans for improving access to climate data sets:

* 1) Development of a catalogue of trusted data sets based on best practices in maturity models, to be used monitoring key climate indicators
* 2) Providing non-technical users with improved access to data. Scope includes finding the trusted datasets, WMO endorsement process, accessibility.

1. The workshop identified information management guidance topics and common practices.
   1. Text

## Summary of workshop

### 1. Introduction to the workshop

1. The WMO Workshop on Information Management was held in Geneva from 2 to 4 October 2017. It was opened by the Chair, M. Jean, President of the Commission for Basic Systems.
2. W. Zhang, Assistant Secretary-General of WMO, welcomed the participants and emphasized the increasing importance of information management practices for members of WMO. He reminded participants that WMO needed to deliver information enabled actions to be taken, especially in the areas of disaster risk reduction and climate. Earth Observation Systems and numerical prediction systems were central to delivering services, but presented challenges in handling the volume of information they could create. WMO members were facing greater volumes of information, with greater diversity, involving more organizations, needing partnerships (both between public sector organizations and between public and private sectors), and opportunities and threats and tools from big data technologies.
3. The aims of the workshop were:

* To identify key climate-related data sources that should be promoted as reference sources and outline a project to enhance their visibility in the WMO Information System to simplify user access to them, and
* to provide guidance on the development of a WMO-wide guide on information management through a review of existing practices and guidance on information management used by WMO Programmes and other organizations.

1. M. Jean, President of the Commission for Basic Systems, introduced the information challenges faced by WMO. He paraphrased Erik Brynjolfsson and Andrew McAfee in “The Second Age Machine”:

*We live in a time of brilliant technologies and the rhythm of innovation is increasing at an unprecedented pace. We are flooded by earth observations, social media provides access to contextual information and unprecedented dissemination mechanisms and high performance computing platform allow us to tackle previously unsolvable problems.* ***It is only a matter of time before the fusion of weather, big data technologies and business applications go mainstream and change the way people and businesses view weather and water data, and experience the force-multiplying effects it will have on improving life and weather sensitive business decisions.*** *Not only is this forcing us to rethink our business models, our recruitment and training strategies and our partnership strategies at the national level, it will also have a fundamental impact on the global meteorological enterprise.* ***If we do not have an information management strategy, we will fail….****.*

1. M. Jean outlined the value chain from data to wisdom, and the stages in managing information from creation to destruction.
2. Summaries of the presentations made during the workshop are in Annex 4.

### 2. Breakout groups to develop the project plan for improving access to climate information

1. The initial view of the breakout teams on defining a set of reference climate datasets was that such a list would be difficult and controversial to maintain. Instead, the groups thought that a method of assigning a level of trust to datasets that was based on objective measures would best help users identify datasets of most relevance to them.
2. Promotion of the most trusted should be through the search mechanisms of specialist catalogues but also the general purpose search engines. The metadata describing the datasets should include reference to the level of trust in a way that the search algorithms could process. With the most trusted datasets appearing high in search rankings, it would be possible to create lists of the most trusted datasets for different application areas; such lists should be linked from the front pages of GISs and WMO web sites.

### 3. Breakout groups to make recommendations on WMO Information Management Guidance

1. Four topics were discussed in breakout sessions: project plan for climate data access; technical support for climate data access, information management guidance topics and, information management common practices.

### 4. Feedback from the breakout groups

#### 4.1 Project plan for climate data access

1. Objectives:

* Global level: authoritative, trusted data sets for informing on key climate indicators (5-10) for global policy users of climate change information. (reference to ongoing work in WMO, GCOS, IPCC)
* Regional and national level: trusted data sets and value added products such as on climate extremes; process for registering data sets in the WIS
* National level: WMO Members commitment for data exchange, how to develop a WIN-WIN approach.

1. Activities:

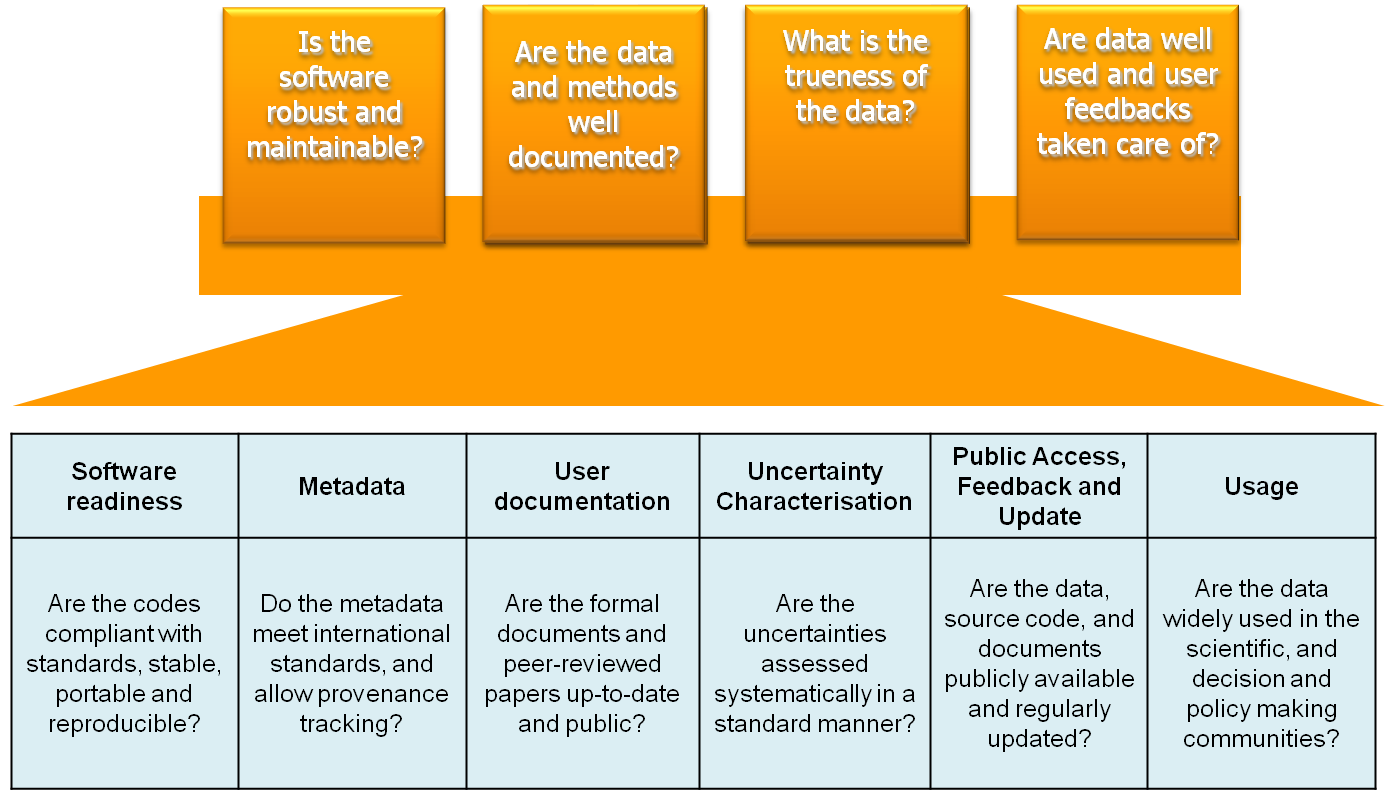
* Assemble inventories / catalogues of relevant and (potentially) trusted data sets
* Adopt a mechanism for assigning data set/system maturity scores
* Adopt a mechanism for assigning application performance scores
* Include (guidance on) uncertainty representation in 1-3
* Make recommendations how to deal with trade-off between NMHS data and voluntary data
* Absorb future developments

1. Assemble inventories / catalogues of relevant and (potentially) trusted data sets.

* Available high quality data sets (both ECVs and derived data/indicators and sources).
* Also include reanalysis, satellite, model data and merged datasets.
* Importance of discovery metadata and simple tools/registries to generate and harvest metadata.
* Encourage each Regional Climate Centre to become a DCPC.
* Describe data license/policy and push for implementation of WMO resolutions.
* Acknowledge cultural differences between weather and climate community with sense of ownership in climate community resulting in restricted data sharing.
* Raise awareness of data sharing and provide something in return, e.g. climate change projections, s2d forecast, etc.
* Liaise with Copernicus C3S for point 1, 2 and 3.
* Recommend open data formats for future sustainable access.

1. Adopt a mechanism for assigning data set/system maturity scores

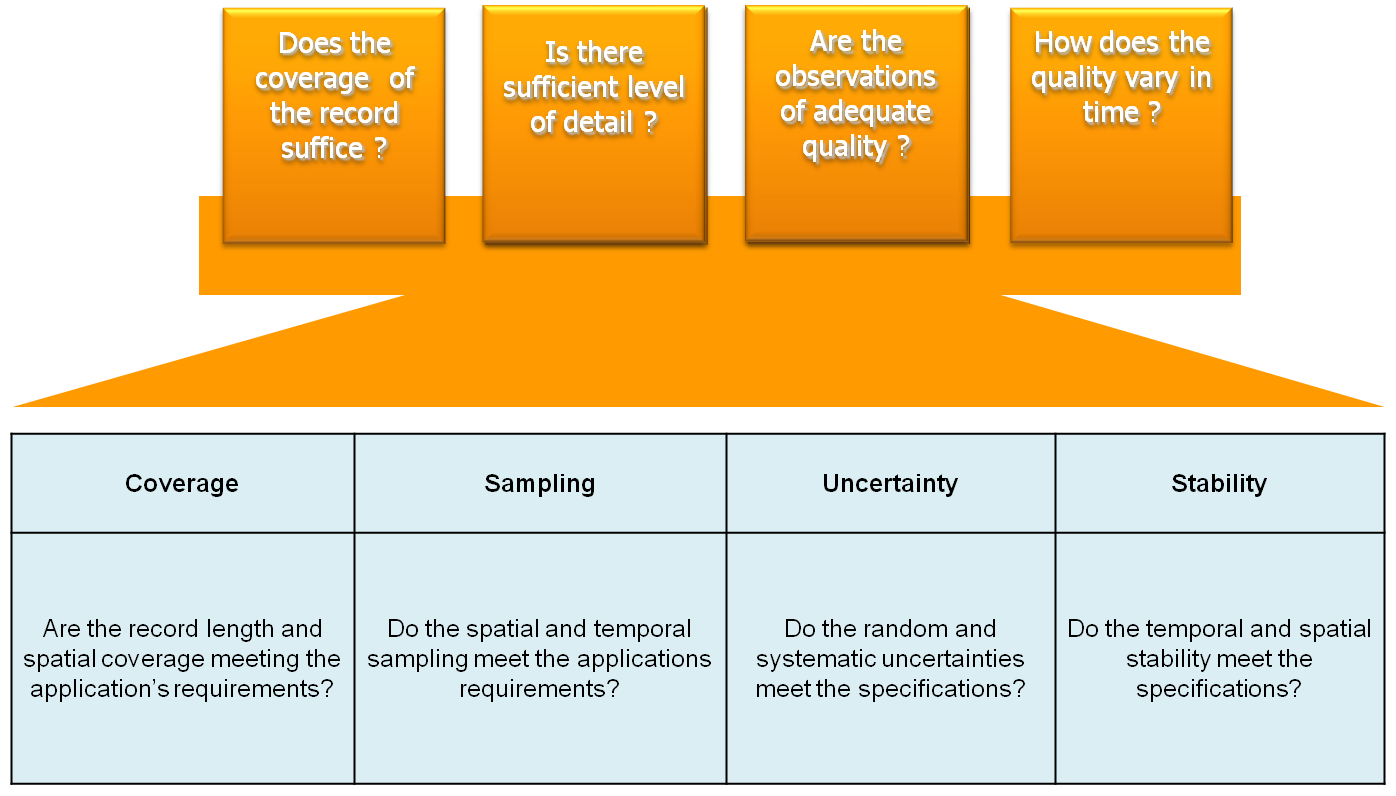
* Available good practices and evaluation studies (such as in EU-Core-Climax)



* Importance of metadata about quality of data.
* Learn from good examples, e.g. WIS, GDPFS for climate data processing and data access similar to the WWW.
* Re-instate the role of IPET-MDRD for guiding on format and coding.
* Define logical data models and involve external users in the identification of data models.
* Identify trusted source institutions for the data sets.

1. Adopt a mechanism for assigning application performance scores

* User application dependent advice



* Categorize data sets based on their performance and utility
* Make plans for communication and management of changes for users including notification mechanism on new versions, decaying ones, etc.

1. Include (guidance on) uncertainty representation.

* Practice of using multiple data sets.
* Develop training material including good practice examples.
* Advise how data gaps issue affect performance in climate monitoring.
* Make recommendations how to deal with trade-off between NMHS data and voluntary data.
* Improved coverage versus lower data quality.
* Address data quality for data from non-NMHSs, such as from farmers, weather amateurs.
* Recommend how lower data quality can be used in NMHS services/operations.

1. Absorb future developments

* Emerging data issues (Internet of Things, Crowd sourcing).

#### 4.2 Project plan for improving access to climate datasets

1. Aim: providing data access to non-technical users. Scope includes finding the trusted datasets, WMO endorsement process, accessibility.
2. Trusted datasets: process endorsed by WMO. Datasets must meet standards: defined by a maturity index approach; Criteria identified the previous day, most met by CORE-Climax and/or NCEI; suitability for purpose; branding of “WMO endorsed”.
3. Architecture considerations. Many requirements met by Copernicus integrated dataset project, which is well-resourced (consider a strategic partnership). Avoid putting all eggs in one basket. Also require governance, common standards etc, and commitment to maintenance. One-stop shop but distributed datasets; this requires interoperability. Use cloud, but have exit strategy. There was a quandary – aim for a limited number of datasets or for a data store with far larger number of datasets? Whichever option is chosen, the representation of the data should conform to a logical data model.
4. Characteristics of interface. Must be intuitive and easy to follow. Reduce infrastructure support needs by reducing complexity and reusing existing architecture. Direct people to WMO priority areas (as per GFCS) – get SMEs to define typical requirements. Users guided through questions. Triage and help desk. Dataset recommendations must meet need (e.g. constraints on climate extremes).
5. Partnerships. With trusted data providers. With DM functionality of RCCs. Important to stress need for individual HMHSs to maintain high standard of observations, DM and specialist product expertise for their areas/ Need infrastructure (CDMS, know-how) at NMHS. Data provision – encourage submission by making effort to demonstrate advantages.
6. Access. One point of entry. Formats should be open and industry standards. Decoded to concentrate at dataset level. Promotion: Google partnership to enhance search-ability; WMO dataset finder”. Communication programme needed. Use experience of other communities to decide (e.g. astronomy), e.g. accessing different datasets.
7. The group emphasized the potential usefulness of using a documented maturity model to select datasets for WMO endorsement. Instead of choosing datasets, which would put the onus on the WMO and could be perceived as "picking winners", a call for datasets could be made. This would create a pool of candidates for endorsement. Those that met the maturity model within a certain range of ranking would be received into a community of endorsed datasets. It is hoped that endorsement would have positive outcomes for the data producers, such as boosting the visibility and overall usefulness of the datasets. It is also hoped that this would be perceived as desirable and encourage other dataset maintainers to improve their data stewardship and seek endorsement of their datasets.
8. It should be noted that such a process would require resources to implement the endorsement process, including periodic review of existing endorsements.

#### 4.3 Information management guidance topics

1. The guidance would need to support NMHS, data publishers, data aggregators, data stewards, data curators and data centres.
2. Guidance would need to support actions to be taken by users and to take into account that not all WMO Members had the same level of skill. It had to be durable; the guidance would need to be valid for many years. An implication was that it could not specify or assume technologies, and would need to refer to external guidance for topics that were subject to rapid change.
3. The intention of WIS Information Management was to change behaviours; this would only be possible if the guidance was not too extensive. Development of a “maturity model” for information management would assist production of guidance that was focussed on the needs of users and users to create action plans to improve their performance.
4. The scope of the guidance should concentrate on the store, share, archive, dispose ^ destroy aspects of the information life cycle. Other sources of information were available for the create and use aspects. There had to be a balance between enabling use of data and protection of data. Although the lifecycle was often written as a linear process, there were often loops within it as datasets were combined or disaggregated. The guidance had to take this into account.

#### 4.3 Information management common practices

1. A summary of which aspects of might need guidance is in [Annex 5](#_Annex_5:_Topics).

### 5. Summary

#### 5.1 Take-away topics for IPET-CDMP

1. W. Wright, chair of IPET-CDMP, outlined the key points arising from the workshop that IPET-CDMP would need to address.
2. The main product of IPET-CDMP would be a high quality global data management framework for climate, including a manual on management of climate data. That manual would address: guiding principles and general requirements, definitions and concepts for climate data, sources of data, guidance on data management for climate, collaborative infrastructure for information management, information management governance, capacity-building and, quality management.
3. Domain-specific practices that would need to be addressed included: quality control and quality assurance, interoperability and data sharing, maturity indices and their value (recommended best practice), meeting user community needs (appropriate datasets), utilising the Cloud, spatial data, partnerships with GFCS priority areas.
4. Cross-domain practices would need to include: communication with users, metadata standards, standard data management practices, data and information policy, DOIs etc, strategy for communicating requirements and recommended best-practice to NMHS and International Data Centre information management, update schedule for guidance, broader approach to communicating and implementing standards, catalogues and discoverability, inventory and catalogue of trusted data, selection of data sets, register of datasets (working with TT-IM).
5. As an interim measure, IPET-CDRM would investigate an objective approach that could be applied to existing datasets to allow a limited number of these datasets to be recognized as the “best” trusted datasets to provide guidance to users before a maturity model approach was used widely enough to be helpful.

#### 5.2 Take-away topics for Task Team on Information Management

1. W. Wright (co-chair TT-IM) summarized the key issues identified for follow-up by the CBS task Team on Information Management (TT-IM).
2. TT-IM would meet for the two days following the workshop; the primary outcomes of the meeting were to be a work plan and a list of the key topics on information management that should be included in the Manual on WIS, together with an indication of the topics on which further guidance would be needed.
3. The provisional time table to be met in order to include information management in the Manual on WIS at eighteenth Congress (Cg-18) was: structure of the additions to the Manual and Guide to the WIS to be presented to OPAG ISS in January 2018; proposed structure discussed at CBS TECO March 2018; EC confirmation that the topic would be considered by Cg-18 in June 2018; consultation with Members on the detailed proposals of for the Manual on Guide during fourth quarter 2018; submission of documents to Cg-18 in March 2019.
4. TT-IM would need to identify: key principles for which standardization would be needed, and would therefore need to concentrate on operations that were common to many application areas; how to ensure that the guidance would be valid for a prolonged period; how to apply maturity model approaches to guide information management improvement activities in Members and to allow “trusted” datasets to be identified; examples of best practice.
5. From the discussions at the workshop, the most likely areas to be given priority were; effective search facilities for discovery of information; interoperability of information management systems; identifying a modest number of preferred representation formats for information; facilitating the combination of datasets of different origin; data preservation, and access to data.

### 6 Workshop close

1. The workshop closed at 1658 on 4 October 2017

\_\_\_\_\_\_\_\_

## Annex 1: Agenda

|  |  |  |
| --- | --- | --- |
|  | **Welcome** | **WMO Assistant Secretary General** |
| **1** | **Information Management** | **President CBS** |
| 1.1 | Keynote: Information Management challenges | M Jean ( President CBS ) |
| 1.2 | Experiences in and new requirements for using climate data | A Klein Tank (Netherlands) |
| 1.3 | WMO Information System – present and future | M dell’Acqua (France) |
| **2** | **Requirements for information management** | **President CBS** |
| 2.1 | Requirements for managing information for climate summaries | O Baddour (WMO) |
| 2.2 | Requirements for managing information for IPCC | Moufouma-Okia (France) |
| 2.3 | A coordinated approach to climate data management: the CCl approach | W Wright (Australia) |
| 2.4 | Global Climate Lead Centres | J Lawrimore (USA) (remote presentation) |
| 2.5 | Requirements for managing information for climate research | M Rixen (WMO) |
| 2.6 | Global Observing System for Climate and Essential Climate Variables | C Lief (USA) (remote presentation) |
| 2.7 | Requirements for managing information for hydrology | T Kanyieke (Uganda) |
| 2.8 | Requirements for managing information for marine meteorology | E Charpentier (WMO) |
| 2.9 | Requirements for managing information for agriculture, meteorology and hydrology, regional perspective | M Waongo (Niger) |
| 2.10 | Requirements for managing information for international civil aviation | D Ivanov (WMO) |
| 2.11 | Information support for the Seamless Global Data Processing and Forecasting System | A Harou (WMO) |
| 2.12 | *Item 2.12 was withdrawn* |  |
| 2.13 | Requirements for managing information for Disaster Risk Reduction | H Bjornsson (Iceland) |
| 2.14 | Information management challenges associated with big data | T Logar (UN GlobalPulse) |
| 2.16 | Requirements for support to the unified service interface | M Androli (WMO) |
| 2.15 | Summary of requirements | Chair |
| **3** | **Improving access to reference climate datasets** | **President CBS** |
| 3.1 | Identifying reference datasets for climate | A Klein Tank (Netherlands) |
| 3.2 | The Copernicus Climate Change Service Global Land and Marine Observations Database: Plans, progress to date, and implications | P Thorne (Ireland) |
| 3.3 | Copernicus/ C3S pulling together climate data, products and services | K Marsh (ECMWF) |
| 3.4 | User interfaces for accessing information | J Tandy (WMO) |
| 3.5 | Hydrohub – providing access to dispersed hydrological information | S Pecora (Italy) |
| **4** | **Project plan – breakout groups (2 on selecting reference datasets, 2 on technical support)** | **Break out group leads** |
| 4.1 | Breakout group reports | Break out group leads |
| **5** | **Identify existing information management practices** | **Matteo dell’Acqua** |
| 5.1 | Guidance for managing large volumes of information | J Peres (France) |
| 5.2 | Guidance for managing international precipitation data | M Ziese (Germany) |
| 5.3 | Guidance for managing climate data and scenarios | R Dunn (UK) |
| 5.4 | Guidance for managing marine information | E Freeman (USA) |
| 5.5 | Guidance for managing national climate observations | B Bannerman (Australia) |
| 5.6 | The Integrated Meteorological Information Service System in CMA: Progress and Outlook | F Gao (China) |
| 5.7 | Use of a maturity matrix to guide management of multi-disciplinary information | C Lief (USA) (remote presentation) |
| 5.8 | Guidance for managing crowd sourced information | H Reges (USA) |
| 5.9 | Guidance for managing international cryosphere information | D Gallaher (USA) |
| 5.11 | ICAO System-Wide Information Management | A Moosakhanian (USA) |
| 5.10 | Chairman’s comments. | Chair |
| **6** | **WMO Information Management Guidance** | **President CBS** |
| 6.1 | Breakout groups (groups A, B, C, D) | Chair |
| 6.1 A | Project plan for climate data access | Klein Tank |
| 6.1 B | Technical support for climate data access | Wiliam Wright |
| 6.1 C | Information management guidance topics | Silvano Pecora |
| 6.1 D | Information management common practices | Matteo dell’Acqua |
| 6.2 | Breakout groups (groups change topics) |  |
| 6.2 D | Project plan for climate data access | Klein Tank |
| 6.2 A | Technical support for climate data access | Wiliam Wright |
| 6.2 B | Information management guidance topics | Silvano Pecora |
| 6.2 C | Information management common practices | Matteo dell’Acqua |
| **7** | **Feedback** | **President CBS** |
| 7.1 | Project plan for climate data access – A & D | Klein Tank |
| 7.2 | Technical support for climate data access – A & B | Wiliam Wright |
| 7.3 | Information management guidance topics – B & C | Silvano Pecora |
| 7.3 | Information management common practices – C & D | Matteo dell’Acqua |
| **8** | **Summary** | **President CBS** |
| 8.1 | Take-away topics for IPET-CDMP | Chair IPET-CDMP |
| 8.2 | Take-away topics for Task Team on Information Management | Chair TT-IM |
| 8.3 | Summary of workshop | Pres CBS |
| 8.4 | Workshop close | Pres CBS |

\_\_\_\_\_\_\_\_\_

## Annex 2: Participants

| **Affiliation** | **Participant** |
| --- | --- |
| AGRYHMET, Niger | Moussa WAONGO |
| Australia | Kate ROBERTS |
| Australia | William WRIGHT |
| Australia | Bruce BANNERMAN |
| Canada | Yves PELLETIER |
| Canada | Michel JEAN |
| China | Yuyu REN (Ms) |
| China | Feng GAO (Ms) |
| ECMWF, UK | Anthony K.P. MARSH |
| France | Wilfran MOUFOUMA-OKIA |
| France | Matteo DELL’ACQUA |
| France | Julie COURIVAUD PERES (Ms) |
| Germany | Markus Johannes ZIESE |
| Germany | Thorsten BUESSELBERG |
| Iceland | Halldor BJORNSSON |
| Ireland | Peter THORNE |
| Italy | Silvano PECORA |
| Japan | Kenji TSUNODA |
| Morocco | Hassan HADDOUCH |
| Netherlands | Albertus M.G. KLEIN TANK |
| Russian Federation | Dmitrii TELIUK |
| Uganda | Tom KANYIKE |
| UK | Robert DUNN |
| USA | Alfred MOOSAKHANIAN |
| USA | Tomaz LOGAR |
| USA | Henry W. REGES |
| USA | Eric FREEMAN |
| USA | David GALLAHER |
| USA | Christina J. LIEF (Ms) |
| USA | Jay LAWRIMORE |
| WMO | Federico GALATI |
| WMO | Peilaing SHI |
| WMO | Omar BADDOUR |
| WMO | Steve FOREMAN |
| WMO | Peer HECHLER |
| WMO | Jeremy TANDY |
| WMO | Michel RIXEN |
| WMO | Dimitar IVANOV |
| WMO | Caterina TASSONE |
| WMO | Simon EGGLESTONE |
| WMO | Ilaria GALLO |
| WMO | Johannes CULLMANN |
| WMO | Fernando BELDA ESPLUGES |

\_\_\_\_\_\_\_\_

## Annex 3: Summaries of presentations

Working documents and presentations supplied for the meeting are available from <http://wis.wmo.int/page=WWIM>.

### 1 Information Management

#### 1.1 Keynote: Information Management challenges (M Jean (President CBS))

1. M. Jean, President of the Commission for Basic Systems, introduced the information challenges faced by WMO. He paraphrased Erik Brynjolfsson and Andrew McAfee in “The Second Age Machine”:

*We live in a time of brilliant technologies and the rhythm of innovation is increasing at an unprecedented pace. We are flooded by earth observations, social media provides access to contextual information and unprecedented dissemination mechanisms and high performance computing platform allow us to tackle previously unsolvable problems.* ***It is only a matter of time before the fusion of weather, big data technologies and business applications go mainstream and change the way people and businesses view weather and water data, and experience the force-multiplying effects it will have on improving life and weather sensitive business decisions.*** *Not only is this forcing us to rethink our business models, our recruitment and training strategies and our partnership strategies at the national level, it will also have a fundamental impact on the global meteorological enterprise.* ***If we do not have an information management strategy, we will fail….****.*

1. In the field of weather forecasting, developments in science, modelling, data assimilation, observing capabilities, and technology had combined to improve the levels of accuracy and detail of forecasts at all time ranges from hours to seasons. These advances come at the cost of increased volumes of data: communications and storage capabilities were not keeping pace with the ability of processors to generate data. As an example, each day the UK Met Office, working on both weather and climate issues, processed 106 million observations, completed 20 quadrillion calculations, archived 10 Terabytes of model data, and created 4 million forecasts.
2. There was a value chain in creating actionable information. At the lowest level are data, such as environmental observations. To be useful, such data have to be transformed to information (such as analyses, forecasts or warnings). Knowledge is interpreted from information to generate, for example, new science (that may be expressed in numerical models) and reports. Knowledge is applied using wisdom to create advice, insight and environmental intelligence.
3. Information passes through several stages during its existence: create, store, share, use archive and destroy. Decisions have to be made for each of these stages, and for some types of information the decision might be made that they should never be destroyed.
4. Information management may be summarized as:

* Deciding what information is needed;
* Making sure the information is fit for purpose;
* Making sure people know about the information;
* Making sure the information is available to (only) those who need it;
* Making sure that the information content is preserved for as long as the information is needed.

1. The key objectives of the workshop were to:

* Identify the main needs of WMO Programmes for help in managing information;
* Identify existing standards, good practice and tools for information management.

#### 1.2 Experiences in and new requirements for using climate data (A Klein Tank (Netherlands))

1. Prof A Klein Tank outlined key aspects in the use of information for climate applications.
2. Climate monitoring is moving beyond assessing global temperature changes to looking at the impacts of climate change on extremes. Between 1910 and 2017 there had been a clear trend for an increase in weather extremes despite inter-annual variability. Studying indices based on extremes allows judgements to be made on whether changes in the characteristics of extreme events may be attributable to man-made influences.
3. Assessing variability was difficult. Many areas had insufficient accessible data to allow a robust analysis; more data may exist, but they were not accessible by the global centres to calculate the variability.
4. Comparing indices around the world was difficult because of the variety of classifications that might be used. The GCOS Essential Climate Variables formed a standard set of indices (27) that had been standardized internationally, so that observations and simulations could be compared in a common way. This allowed trends in different areas to be compared, and simulations in different areas to be compared with each other and with observations.
5. A seamless approach to climate services would need to provide advice on past climate (climate references), trends in current weather event attribution, weather forecasting, seasonal forecasting and climate scenarios.
6. Historically, when looking at averages, it had been adequate to look at monthly averages. However, with a focus was on extremes it is necessary to look at observations throughout the day, which means going back to archives that, even in advanced countries, may not have been digitized.

#### 1.3 WMO Information System – present and future (M dell'Acqua (France))

1. M. dell’Acqua (France) described the WMO Information System (WIS).
2. WIS was about finding, accessing, exchanging and managing information (Parts A, B and C). Part A was an evolution of the Global Telecommunication System for routine exchange of information, Part B extended this to discovery, access and retrieval of information; Part C was being built and was concerned with how information is managed.
3. Anyone could use the WIS to access information, but some of the information was limited to registered users because of use constraints placed on the information by ownders of the information.
4. Within WIS, National Centres were responsible for gathering or creating information nationally and making agreed information available to the international community, and for receiving information from the international community and making it available nationally. Data Collection or Production Centres were responsible for creating, collecting and distributing information to meet an international requirement of a programme of WMO or partner organization. Global Information System Centres held the catalogue of information that was being published through the WIS and managed the international flow of information that was exchanged routinely.
5. WIS encourages the use of standard data formats, but was built to use any data format needed by users.
6. WIS 2.0 sought a way of better meeting the needs of users from across all WMO and partner programmes and that would cope with the trends in technology and data volumes (see <http://wis.wmo.int/file=3544>). It was intended to provide a “virtual one-stop-shop” for weather, water and climate information and services. To do this it would provide an environment in which data could be: managed, documented, discoverable, accessible and easy to use. It would also standardize information management so data could be relied on.
7. Questions focussed on the challenges of data rescue, that was not only concerned with recovering data from paper, but also with ensuring that information on old digital media could be accessed in the future. Not only were the means of reading media becoming obsolete, but the software needed to interpret the data was often not available.
8. During discussion it was suggested that information should be stored in open formats were needed (rather than instrument proprietary format) to make it possible (in the future)to read old data. As data stores became more complex, it was becoming more difficult to extract the required information – for example different sites might require different versions of Java. Direct connection to the data might remove some of the barriers to reading different formats. M. dell’Acqua commented that the approach of WIS was not to build-in a standard format, but that there was a strong argument for providing standard (but perhaps inefficient) interfaces that gave easy access to the information content.
9. W Zhang (ASG) commented that data management and data policy went hand in hand. Without a data policy of exchange, there was little point in using international management standards. The World Weather Watch had maintained a small rate of increase in the volume of stations exchanging data. As we moved to higher resolution forecasting, more observations would be needed for initial conditions and validations. Over the past 30 years there had been an increase in remote sensing data – but the large volumes and complexity of data structure (and hence formats) provided challenges that needed the providers and users to cooperate in order to achieve a level of data exchange that met the needs in a way that was affordable and possible (from a telecommunications perspective). Data policy for climate (Res 60) had not led to a significant increase in data availability – more effort was needed to improve the availability of data.

### 2 Requirements for information management

#### 2.1 Requirements for managing information for climate summaries (O Baddour (WMO))

1. O. Badour (WMO) summarized the key information management requirements to support production and publication of climate summaries. Whereas in the past climate summaries had concentrated on average conditions, emphasis was moving towards publication of statistics on daily extremes and widespread extreme events.
2. Globally coordinated climate summaries were intended to:

* Provide snap-shots of the state of the climate on global, regional, national space scales, and on monthly, seasonal, annual and multi-year time scales;
* Ensure authoritative summaries that were based on the best available data, provided peer-reviewed analysis of those data, include products whose quality had been checked and that included estimates of uncertainty;
* Complement the longer update cycle of the IPCC Climate Assessment process;
* Provide special assessment reports when required.

1. Examples of climate summary publications included: annual statements on the state of the global climate that have been published since 1993, (World Meteorological Organization, 2016), global climate evolution during the periods 2011-2015 (World Meteorological Organization, 2016) and 2001-2010 (World Meteorological Organization, 2013), and special reports on extreme conditions in the northern hemisphere published in 2010 (World Meteorological Organization, 2010) and 2012 (World Meteorological Organization, Japan Meteorological Agency, Deutscher Wetterdienst, 2012).
2. The WMO global statements on the stat of the global climate included information on: global temperature (globally and by continent), the cryosphere, global precipitation, sea levels, greenhouse gases, extreme events, and the impact of events.
3. Production of climate summaries relied on contributions from countries being based on consistent methodologies for reporting, having geographical balance and representativeness, and on those contributions having been quality checked and verifiable. The creation of the global datasets built on these contributions were based on peer-reviewed methodologies, recorded the traceability from the input data to the output datasets, were produced in a timely manner and had to be easily accessible by users.
4. In addition to the World Weather Records that had been produced since 1927, climate data were needed to support calculations of climatological standard normals, daily climate extremes, national climate monitoring products, and for reporting wide-=area extreme events.
5. Traditional methods of estimating the daily extreme temperatures from synoptic reports had proven to be inadequate for climate purposes. To address this, a new specification for reporting daily climate information using the table driven code forms had been prepared. That specification allows 31 daily observations to be included in a report that contains: time of observation for temperature, daily maximum temperature, daily minimum temperature, daily mean temperature (if it differs from the average of the maximum and minimum temperatures), time of observation for precipitation, total daily precipitation, depth of new snowfall and depth of total snow on the ground.
6. The requirements for information management to support these activities are summarized in the following bullet points.

##### Availability

* Friendly gateway(s) for discovering data;
* Provide information on inventory of available data on essential Climate Variables (ECVs).
* Information on data rescued and still to be rescued (Data Rescue information management).
* National Climate data and products to be reported using commonly recommended practices.

##### Exchange

* Information on data exchange policy.
* Information on data model and format.
* Date of next update if dataset (if regular updates).
* Guidance on Access and Retrieval.
* Guidance on the registry of data sets and products.

##### Use

* Users informed about publication of new data sets and versions.
* Description of data sets for global use: information on the peer review, origin/ author.
* Guidance on maturity and uncertainty estimate.
* Guidance on how to cite datasets.
* Guidance on software for analysing data.

#### 2.2 Requirements for managing information for IPCC (Moufouma-Okia (France))

1. W. Moufouma-Okia (France) presented the information management needs in support of the work of the IPCC.
2. The role of the IPCC was “… to assess on a comprehensive, objective, open and transparent basis the scientific, technical and socio-economic information relevant to understanding the scientific basis of risk of human-induced climate change, its potential impacts and options for adaptation and mitigation.” It was structured into three Working Groups and one task force. The presentation focussed on the work of Working Group I (WGI) – the physical science basis. IPCC recommendations were intended to be “policy-relevant but not policy-prescriptive.”
3. Since is foundation in 1988, the IPCC has produced peer-reviewed reports on the science of climate change. In addition to wining the Nobel Peace Prize in 2007, it had provided key input that led to the Kyoto Protocol in 1995 and the Paris Agreement in 2014.
4. To meet its aim of “reliable detection and attribution of changes in climate, and their effects, is fundamental to our understanding of the scientific basis of climate change, WGI applied working principles of rigour, scientific accuracy, and comprehensiveness, traceability of findings, transparency of findings, use of peer-reviewed literature, consistent treatment of uncertainties, reconciling assessment of multiple data products, assessing performance of multi-model climate simulations, and combining multi model climate projections.
5. The key lessons from the work of the IPCC were:

* There was a need to ensure a transparent and traceable account of IPCC key findings including data sources used to provide multiple lines of evidences and figures.
* There was a need to ensure curation of data sources.
* Archiving observational datasets and model data outputs was beyond the current capacity of the IPCC Data Distribution Centre – WMO programmes (GCOS, WCRP etc) could play a crucial role to support this requirement.
* IPCC WGI Technical Support Unit could help compile “recipes” for producing figures of information contained in Working Group I reports.
* The IPCC data distribution centre hosted the reference archives for climate data outputs that underpinned the IPCC assessments. It was designed to facilitate timely distribution of consistent scenarios of changes in climate and related environmental and socio-economic factors for use in climate impact and adaptation assessment. It ensured that unambiguous attribution could be given to authors and distributors, data curators and institutions responsible for reviewing or otherwise ensuring the quality of the resources it held.

#### 2.3 A coordinated approach to climate data management: the CCl approach (W Wright (Australia))

1. W. Wright (Australia) described the approach to climate data management being taken by the Commission for Climatology.
2. Key aspects of the requirements were addressed in the Climate Data Management Specifications (World Meteorological Organization, 2014). At the time of the workshop, none of the 96 known climate data management systems fully met the specifications.
3. High quality climate datasets were required to provide authoritative climate assessments, ensure the integrity of climate data, to develop climate change knowledge to inform adaptation strategies at global, regional and national levels, and to support climate modelling and climate change projections.
4. Specific requirements for climate datasets were that they should be: as long a time series as possible, accessible in digital forms, as far as possible complete, properly quality controlled, have confidence assessments, as far as possible free of spurious discontinuities and trends (requiring attention to observing practices), and accompanied by adequate metadata and complete documentation of lineage.
5. The Commission for Climatology was planning a High Quality Global Data Management Framework for Climate (HQ-GDMFC) that would include a manual on climate data management. The need for such a framework was driven by the following factors. Much existing guidance on climate data management was out of date, due, for example, to rapid recent advances in technologies. The NMHS of many countries reported lack of capacity in (climate) data management. There was a lack of a regulatory framework, standardisation of terminology, processes and policies. The framework was also needed to enable better use of new data sources and advances in technology.
6. Initial findings of the team responsible for creating the HQ-GDMFC (IPET-CDMP – the Inter-Programme Expert Team on the Climate Data Modernization Programme) were that: there was scope for collaborative arrangements for data management between different scales and domains (international datacentres, regional climate centres, National Meteorological and Hydrological Services (NMHS) and the WIS. The team also concluded that there was a need for new standards on management of changes to observing systems and processes, process documentation, and authentication and publication of climate datasets.

#### 2.4 Global Climate Lead Centres (J Lawrimore (USA) (remote presentation))

1. J. Lawrimore (USA) gave a remote presentation on data management at NOAA’s National Centres for Environmental Information(NCEI) as an example of how Global Climate Lead Centres were tackling the issues of information management.
2. NCEI used a model of six tiers data stewardship to assess the appropriate level of handling for each dataset and whether NCEI was fulfilling its responsibilities effectively:  
   1) Long-term preservation and basic access,  
   2) Enhanced access and basic quality assurance,  
   3) Scientific improvements,  
   4) Derived products,  
   5) Authoritative records,  
   6) National services and international leadership.

##### Level 1 – long-term preservation and basic access

* Preserve data with metadata adequate for basic discovery and DOI (Digital Object identifier) minting,
* Provide basic access to the original data and ensure the safeguarding of the dataset over its entire life cycle,
* Serve as expert advisors on data formats, conventions, metadata creation, and standards for data providers.
* *NCEI examples: Comprehensive Large Array-data Stewardship System (CLASS) Satellite Archive, Meteorological Assimilation Data Ingest System (MADIS).*

##### Level 2 – Enhanced access and basic quality assurance

* Provide enhanced data access through specialized software services for users and applications,
* Support users’ ability to conduct quality assurance, sub-setting, and statistics collection.
* *NCEI examples:* *National Operational Model Archive and Distribution System (NOMADS), THREDDS Data Servers, Climate Data Online (CDO).*

##### Level 3 – Scientific improvements

* Improve data quality or accuracy with scientific quality assessments, controls, warning flags, and corrections,
* Re-process datasets to new, improved versions or formats and distribute to users.
* *NCEI examples: Global Temperature and Salinity Profile Program (GTSPP), NEXRAD Multi-Radar/Multi-Sensor Reprocessing (2002-2011), Processed 1-minute DART and tide data for tsunami research.*

##### Level 4 - Derived products

* Build upon archived data to create new products that are more broadly useful,
* Distil, combine, or analyse products and data sources to create new or blended scientific data sets and products.
* *NCEI examples: International Best Track Archive for Climate Stewardship (IBTrACS), Integrated Surface Dataset (ISD), International Comprehensive Ocean-Atmosphere Data Set (ICOADS), VIIRS & DMSP Night-time Light and Fire Products.*

##### Level 5 – Authoritative records

* Combine multiple sources into a single, inter-calibrated data product,
* Authoritatively establish the quality and provenance for a data product,
* Ensure the product is fully documented, reproducible, and its production methods are transparent by preserving the data, metadata, algorithms, and workflows used in creating the product.
* *NCEI examples: Global Historical Climatology Network (GHCN), International Satellite Cloud Climatology Project (ISCCP), Other Climate Data Records (CDRs).*

##### Level 6 – National services and international leadership

* Lead, coordinate, or implement scientific stewardship activities for a community or across disciplines,
* Provide highly specialized levels of data services and product assessments based on information products and scientific knowledge.
* *Examples: Climate Monitoring/State of the Climate, National Climate Assessment (NCA), Intergovernmental Panel on Climate Change (IPCC) , National Integrated Drought Information System (NIDIS).*

1. Specific issues that needed to be addressed when manging datasets at higher tier levels included: building and maintaining historical data holdings, maintaining datasets through operational ingest, maintaining datasets through other means, building a dataset from multiple sources, quality control, metadata, data and product access.
2. Some datasets had been developed using a variety of sources acquired through various means including:

* Direct agreements with other meteorological and data centres that collect, archive, manage and distribute climate data (for example the Australian Bureau of Meteorology had provided NCEI with daily data for 17,000 stations), but data sharing restrictions within some countries limited data access;
* Related projects (e.g., WMO normals, World Weather Records, Colonial Era Archives), the need to develop personal contacts – the World Weather Records (Decadal/Annual data collections) was an excellent source of data, and the greatest success in collecting data from countries was often in association with a request form WMO, but the processing and conversion of multiple and diverse formats was time consuming;
* Personal contacts; and
* International Workshops.

1. Ongoing updates to datasets were accomplished largely through operational automated acquisition and ingest processes, including: GTS transmissions, monthly means (via CLIMAT bulletins), synoptic reports (observations every 3 or 6 hours, that could include daily maximum and minimum temperatures), METAR (hourly observations, from airports), web services (countries such as Canada had provided data daily via web services), file transfer protocol (ftp), and a small amount of data was still being received by e-mail.
2. When building datasets from multiple sources, particular issues that needed to be addressed include:

* Formatting: converting between formats can be labour intensive; manual effort is needed to decipher and align or remove extraneous lines, columns and data;
* Maintaining metadata: ensuring that the inventory contains all the stations in the data file;
* Duplicates: there are many different ways to calculate monthly mean temperature, and different data sources often use similar (but not exactly the same) method and it is often not possible to determine a single definitive time series;
* Metadata: inaccuracies complicate merging of station records - even if large percentages of data matched, large distances between stations made station matching difficult;
* Other complicating factors: changes of identifiers that occurred that were unrelated to station moves or stations that are members of multiple networks and consequently had multiple identifiers;
* Identifying corresponding stations in different data sources: comparisons of overlapping observational data between sources were often required to match stations and there were rarely exact data matches.

1. NCEI relied heavily on automated, objective and repeatable processes for quality control of monthly, daily and sub-daily data. As new algorithms were developed, entire periods of data records could have the new quality criteria applied. Manual corrections to data were not made operationally, but there was a mechanism for incorporating manual changes retrospectively and NCEI used a web-based reporting mechanism so that external users could report possible data errors.
2. Accurate metadata were essential for proper data management. Metadata often consisted only of station location and elevation; detailed station histories had yet to be exchanged internationally, and much of the global metadata used by NCEI was outdated. There was little or no information available about instruments, practices or the physical environment surrounding the observations. Some metadata could be inferred from sources of information such as digital elevation models, night-light data from satellites (used to distinguish between rural and urban areas) and imagery datasets (such as the maps provided by major internet services).
3. NCEI provided many ways to access data and products: bulk download (using ftp or https), web service API (server to server downloads), web search and discovery services (OneStop; https://www.ncdc.noaa.gov/onestop/#/ ), web order fulfillment services (Common Access) and enterprise application services including THREDDS (easy downloads of NetCDF data) and ERDDAP (download subsets in common formats and make graphs & maps).

#### 2.5 Requirements for managing information for climate research (M Rixen (WMO))

1. M. Rixen (WMO)summarized the needs of the World Climate Research Programme’s (WCRP) activities for information management.
2. The mission of the WCRP was “… to facilitate analysis and prediction of Earth system variability and change for use in an increasing range of practical applications of direct relevance, benefit and value to society”. Its two overarching objectives ere to determine the predictability of climate and the effect of human activities on climate.
3. The WCRP Coupled Model Inter-comparison Project (CMIP) defined common experiment protocols, forcing and output to allow different coupled climate models to be compared.
4. The WCRP Coordinated Regional climate Downscaling Experiment (CORDEX) sought to advance the science and application of regional climate downscaling for improved regional climate information.
5. WCRP had a vision for a data infrastructure known as the Earth System Grid Federation. This would share observations, reanalyses and simulations together with estimates of their uncertainty. It already held 701 244 datasets totalling 4 636 TB.
6. WCRP needs were for: techniques to manage and use “big data”, a distributed infrastructure, an open data policy, metadata and documentation of data, interoperability of data sources, a maturity index for datasets, uncertainty or confidence level of data, guidance on how to use data, accountability for data, versioning of datasets, governance of data management, a “disclaimer” to address liability issues, co-design of research and operational activities, user feedback, innovation at all stages in data creation and use, observations that were sustained and calibrated against international standards, and adequate funding for data related activities.

#### 2.6 Global Observing System for Climate and Essential Climate Variables (C Lief (USA) (remote presentation))

1. C. de Groot-Lief (USA) described the GCOS Essential Climate Variable data access and information matrix that was intended to support easy access to high quality global datasets.
2. The Global Observing Systems Information Center (GOSIC) portal ([https://GOSIC.org](https://gosic.org/)) hosted the Global Observing System for Climate (GCOS) ECV Data Access and Information Matrix. This provided access to:  
   **Atmosphere ECVs**: Surface: Precipitation, Pressure, Surface Radiation Budget, Surface Wind Speed and Direction, Temperature, Water Vapour;  
   **Upper Atmosphere**: Earth Radiation Budget, Lightning, Temperature, Water Vapour, Wind speed and Direction;  
   **Atmospheric Composition**: Aerosols properties, Carbon Dioxide, Methane and other Greenhouse gases, Cloud Properties, Ozone, Precursors (supporting the Aerosol and Ozone ECVs);  
   **Ocean ECVs - Physical**: Ocean Surface Heat Flux, Sea Ice, Sea Level, Sea State, Sea Surface Salinity, Sea Surface Temperature, Subsurface Currents, Subsurface Salinity, Subsurface Temperature, Surface Currents, Surface Stress;  
   **Ocean ECVs - Biogeochemical**: Inorganic Carbon, Nitrous Oxide, Nutrients, Ocean Colour, Oxygen, Transient Tracers;  
   **Ocean ECVs - Biological/Ecosystems** : Marine Habitat Properties, Plankton;  
   **Land ECVs**: Above-ground biomass, Albedo, Anthropogenic Greenhouse Gas Fluxes, Anthropogenic Water Use, Fire, Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Glaciers, Groundwater, Ice Sheets and Ice Shelves, Lakes, Land Cover, Land Surface Temperature, Latent and Sensible Heat fluxes, Leaf Area Index (LAI), Permafrost, River Discharge, Snow, Soil Carbon, Soil Moisture
3. The top-level interface to the matrix provides a one-page table of links, with one link for each of the ECVs. Those links lead to a more detailed page on the ECV, from which there are links to pages describing and giving access to individual datasets. A dashboard also shows summary trends of key climate information.
4. For users to efficiently discover and access datasets, the experience of NCEI was that the following were required:

* Archival of data using submission agreements (so that users follow an end-to-end process standard operating procedure);
* Standardized data documentation (metadata). NCEI used the ISO 19115-2 standard and standardized key words from the NASA Global Change Master Directory (GCMD) as well as standard data access formats. Metadata was displayed using a standardized landing page template;
* Assigned subject matter experts (SME) were in charge of their data sets and documentation generation and maintenance. NCEI had a paradigm shift in 2014 and trained SMEs to be in charge of their own data documentation. This was accomplished by providing easy to use text based tools to generate standardized metadata records (one such tool NCEI uses is ATRAC: https://www.ncdc.noaa.gov/atrac/index.html);
* Minting DOIs when appropriate;
* Making data available through general data search and specialized data access tools such as matrices or clearinghouses. Users wanted to be able to go to one location as well as find a data collection such as the ECVs.

#### 2.7 Requirements for managing information for hydrology (T Kanyieke (Uganda))

1. T. Kanyieke (Uganda) described the information requirements of the hydrological community.
2. Hydrological information was used for: assessing a country’s water resources; planning; designing, and operating water projects; assessing the environmental, economic, and social impacts of water resource management practices; assessing the impacts on water resources of other non-water sector activities, such as urbanization or forest harvesting; providing security for people and property against water-related hazards, particularly floods and droughts. (World Meteorological Organization, 1994).
3. Requirements for hydrological information fell into two categories: to sustain programs and projects; and to support informed Decision Making in Water Resources Management. (Mosely, 2001).
4. The key elements of a hydrological information system were: network design, data management, quality management, training and technology, information necessary for the provision of services in support of the protection of life and property and well-being of all nations.
5. To address the issue of hydrological data being focused on river basins and local catchment areas rather than on a global dimension it was necessary to support collaboration to establish linkages from catchment to national level, to take advantage of existing linkages between national level and transboundary regional levels, and to scale-up the transboundary regional levels to global.
6. To address the issue of hydrological data being focused on river basins and local catchment areas rather than on a global dimension, there was a need to support collaboration:

* Establishment of linkages from catchment to national level;
* Taking advantage of existing linkages between national level and transboundary regional levels;
* Scaling up the transboundary regional levels to global;
* Link to the hydrological information service from the NMHs
* Need to set-up national Hydrology Information Systems with the following attributes:
  + well-designed operational & maintained network,
  + selection of appropriate & compatible technology,
  + system for data management was maintained,
  + quality management,
  + a training program for continuity.

#### 2.8 Requirements for managing information for marine meteorology (E Charpentier (WMO))

1. E. Charpentier (WMO) outlined the requirements of marine meteorology for managing information.
2. The Joint Commission on Oceanography and Maring Meteorology (JCOMM) gathered observations: upper ocean temperature and salinity, surface underway temperature and salinity, underway surface ocean and marine meteorological data (SAMOS),Argo profiling float data, drifters, meteorological moored buoys, tropical moored buoys, tsunameters, ocean reference sites, voluntary observing ships, tide gauges, sub-surface gliders, high resolution sea surface temperature data from satellites.
3. Information management requirements for marine meteorology were centred on a need to review the draft Joint WMO and IOC Strategy for Marine Meteorological and Oceanographic Data Management (2018-2021) and determine how CBS can contribute to it; feedback to be provided to the JCOMM Data Management Coordination Group. In particular, the following topics should be addressed:

* Outcome 3, Activity 1 : Updating Cookbook for Submitting Data in Real Time and in Delayed Mode;
* Outcome 3, Activities 2 and 3, Outcome 5, Activity 1: Assisting for making ocean data centres interoperable with the WIS (ODP, ICOADS, WOD, CMOCs, GDACs, Copernicus …);
* Outcome 3, Activity 4 : Assist with regard to the JCOMM OpenGTS Project;
* Outcome 3, Activity 6: Provide information on relevant centres which are documenting their QC/QA procedures;
* Outcome 4, Activity 1 : Make relevant WMO standards visible via the OceanBestPractices repository;
* Outcome 4, Activity 2 : Assist for update of existing and development of new BUFR templates (gliders, rigs & platforms, BGC, HF radars…);
* Outcome 5, Activity 2: Contribute to development of the future IOC ODIS.

##### Outcome 6, Activity 1: WIS training activities to consider ocean data and information issues in its curriculum.

1. Additional information was available from (Intergovernmental Oceanographic Commission, 2017), (World Meteorological Organization, 2017), (World Meteorological Organization and Intergovernmental Oceanographic Commission, 2017), (World Meteorological Organization and Intergovernmental Oceanographic Commission, 2017),

#### 2.9 Requirements for managing information for agriculture, meteorology and hydrology, regional perspective (M Waongo (Niger))

1. M. Waongo (Niger) described the AGRHYMET Regional Centre (ARC) of the Comitè Permanent Inter-États de lute controle la sécheresse dans le Sahel as an example of a regional perspective on the information needs in support of services for agriculuture, meteorology and hydrology.
2. AGRHYMET products were built on 700 rain gauges, of which 300 were used in near-real time for agricultural monitoring (and of those 80 were at synoptic stations). These were complemented by approximately 250 hydrometric stations.
3. Availability of records of observations appropriate for climate use varied greatly from country to country. Temperature and precipitation records for Guinea were available from the mid1890s to the present day, but precipitation records for Ghana were only available from 1960, and temperature records from Côte d’Ivoire from 1989. Delays in processing observations for climate were uneven, with precipitation data for Cabo Verde only available until 2011, and temperature records from Sénegal until 1990.
4. ARC used a data sharing policy. Climate data were made available free of charge for scientific purposes (research, education). ARC did not have the right to share data they hosted on behalf of a country without explicit agreement from that country. ARC staff were allowed to use the hosted climate data only after signing a user’s charter. Use of the data by others was subject to case-by-case agreement with countries. A Memorandum of Understanding on sharing of climate data that would simplify the procedures for supplying data had been created in 2011 but had not been signed by all contributing countries.
5. Products generated by AARC were distributed to end users who included policy makers, on-governmental organizations, private sector companies and farmers. Products included 10-day, monthly, seasonal and special bulletins. Distribution was through printed documents, radio and television broadcasts and the website [www.agrhymet.ne](http://www.agrhymet.ne).
6. AGRHYMET saw the most urgent needs related to information management to be:

* Enhanced infrastructure (hardware, software, internet) and human capacities to collect data, produce, organize and effectively deliver climate products at regional level;
* Regional Climate Centres in the Economic Community of West African States deliver effective services;
* Effective Memoranda of Understanding on data sharing, with the help WMO to speed up the process with the NMHSs;
* Optimization of the regional climate observation network;
* Improved data quality control at regional level (especially removing errors when combining data from different sources);
* Enhanced capacity of NMHSs or data collection; and
* Improved storage, analysis and dissemination of climate data.

1. During questions it was commented that only monthly climate data were available before 1960 and more detailed information could not be recalculated because the more detailed working documents that had been used to create the monthly data had been lost.

#### 2.10 Requirements for managing information for international civil aviation (D Ivanov (WMO))

1. D. Ivanov (WMO) outlined the requirements of international air navigation for meteorological information management.
2. Aeronautical meteorological information was: highly regulated for context, format, exchange procedures and quality) at international and national levels; designed to be fit-for-purpose based on concrete user requirements; well-defined and with a recognized purpose (safety, regularity and efficiency of air transport operations); in support of decisions; for restricted use and; provided through cost-recoverable services.
3. The regulations governing meteorological informatoin in support of international air navigation were defined in information (International Civil Aviation Organization, 2015) that was also published as (World Meteorological Organization, 2016).
4. The regulatory framework covered: governance, roles, responsibilities; establishment of international information services and facilities – WAFS (WAFCs), IAVW (VAACs), TCACs; Standardization of observations and reports, forecasting and warning procedures, aeronautical climatological information, use cases for services to stakeholders; flight documentation ; telecommunications and information exchange.
5. ICAO regulations required that “each Contracting State shall designate the authority, hereinafter referred to as the meteorological authority, to provide or to arrange for the provision of meteorological service for international air navigation on its behalf.” This separated the roles of specifying the information required with appropriate quality procedures, and providing the information.
6. Operational information (OPMET) was delivered in alphanumeric form (Traditional Alphanumeric codes) for observations (METAR, SPECI, local reports and aircraft reports), forecasts (TAFs), advisories and warnings (SIGMET, AIRMET, Volcanic Ash and Trobical Cyclone advisories, aerodrome warnings, significant weather advisories), in binary form of GRB (World Area Forecast System forecasts for wind, temperature, icing, turbulence and cumulonimbus) or BFR (WAFS significant weather forecasts), or in graphical form (significant weather charts, graphical advisories, graphical SIGMET.
7. The following requirements were identified:

* Recognize the high level of regulation and standardization at international level (ICAO and WMO) of the aeronautical meteorological information (International Civil Aviation Organization, 2015) also published as (World Meteorological Organization, 2016);
* Consider the requirements for quality management of the aeronautical meteorological information raised at the level of a standard as a good practice;
* Note the ICAO process for establishing user requirements and consider its applicability within the WMO information management;
* New requirements for the MET information stemmed from the need for further integration in the stakeholders’ decision-making (mostly automated) systems and processes;
* There was a strong user requirement for more near-real-time information uplinked directly to the cockpit to support near-term decisions (e.g., 0 to 20 min decision horizon);
* There was a need to maintain close coordination and cooperation between ICAO and WMO in the information management domain; interoperability between WIS and SWIM (International Civil Aviation Organization, 2014) was a concern for some information providers;
* Aviation users were getting more interested in potential climate change impact on aviation – there was a need to develop appropriate ways to present such information to various stakeholders.

1. Trends and developments in the aeronautical meteorology with regard to the information management, as stipulated in the ICAO Global Air Navigation Plan (GANP, (International Civil Aviation Organization, 2016)) and its Aviation System Block Upgrades (ASBU) methodology included:

* Transition from TAC forms to IWXXM;
* Integration of MET information in the Air Traffic Management (ATM) decision-making systems;
* The migration from a “product-centric” to a “data-centric” information service;
* Implementation of the ICAO System-Wide Information Management (SWIM) concept;
* Further regionalization and globalization of service provision.

#### 2.11 Information support for the Seamless Global Data Processing and Forecasting System (A Harou (WMO))

1. A. Harou (WMO) summarized the Global Data Processing and Forecasting System (GDPFS) and how it used information. The Seamless Global Data Processing and Forecasting System was expected to extend the GDPFS to provide a framework for the routine production to support all WMO programmes in the same way that WIGOS and WIS had expended beyond support for real time weather forecasting. It would concentrate on developing and delivering standard products upon which sector-specific services could be built. It would also facilitate the creation of impact-based forecasts and risk-based warnings.
2. Impact-based forecasts and risk-based warnings introduced new challenges: integrating non-conventional data, such as exposure and vulnerability to weather, water and climate events, would need new approaches to managing information, much of which would not come from NMHSs.
3. The GDPFS used data representations in the Manual on Codes and the exchange facilities of the WIS to deliver its traditional products; It was also introducing audits of GDPFS centres as a means of assuring the quality of its products. New and enhanced information management practices were most likely to be needed to support impact-based forecasting and risk-based warnings.

#### 2.13 Requirements for managing information for Disaster Risk Reduction (H Bjornsson (Iceland))

1. H. Bjornsson (Iceland) described hazard monitoring as practised at the Icelandic Meteorological Office (IMO), different categories of operational hazard management, and how effective information management could assist in the aim of reducing the impact of disasters.
2. Disaster management activities were often discussed using a “disaster management cycle” consisting of preparedness, response, recovery and mitigation. The Icelandic Meteorological Office was mainly concerned with preparedness and response. It was particularly involved in monitoring, analysing, interpreting, informing, giving advice and counsel, providing warnings and forecasts and where possible, predicting natural processes and natural hazards, and in issuing public and aviation alerts about impending natural hazards, such as volcanic ash, extreme weather and flooding.
3. Supporting its disaster management activities were over 600 instruments used to observe natural hazards and the long term behaviour of geophysical systems. IMO was highly reliant on remote sensing products and numerical weather prediction (WP) models to support its disaster management activities. The forecast model that covered both Iceland and Greenland was run jointly by IMO and the Danish Meteorological Institute.
4. Although volcanic ash was a major hazard from volcanic activity (the eruption of Holuhraun in 2014 produced a 85km2 lava field in a six month period – equivalent to covering Manhattan Island to a depth of five floors, the production of SO2 (at a rate of 20-40 kT a day) represented a major hazard to health. Trajectory modelling was a key component of managing the incident. Although dangerous, the spread of the gas plume was difficult to monitor directly, so a crowd sourcing method was used whereby citizens could use a web page to provide information about smelling gas or experiencing irritation that might have been caused by a gas plume.
5. During the preparedness phase of an event, relevant data needed to be gathered and analysed. These data should be standardized and available over a long period. The data also have to be relevant (as opposed to easy to measure) – snow accumulation at a weather station may not be representative of the accumulation at locations at which avalanches are triggered.
6. During the lead-up to an event, information on the precursors or triggers of an event become critical. (a rainfall amount with return period greater than 30 years was likely to cause flooding events; some volcanoes exhibit characteristic activity in the period leading up to an eruption). At this stage, rapid and clear exchange of information between authorities becomes important and needs to use pre-agreed standards.
7. During an event there is no time to develop new sources of data. This is a particular challenge, because information systems are built on the lessons from the past, and there is a natural tendency for systems to process the information that would have helped manage the previous event – rather than the one taking place. The formats of data also had to be agreed well in advance of the data being used.
8. Observational meteorology had more standardization than other branches of geoscience, but even so there were often uncertainties over what had been reported (for example the sampling period for wind gusts). Time and effort was needed to develop relationships with practitioners in other disciplines to ensure effective data exchange and use.
9. IMO used some publicly sourced data, such as ash fall, spread of gas of volcanic origin. Although useful for both IMO and the public, the data were of varying quality. Citizen-sourced data was a useful source of supplementary information, but could not be relied on as the primary means of observing.
10. The key conclusions were:

* Risk reduction depends on good data for preparedness and for an effective response during an event;
* Good data must be relevant, based on good measurements, preferably from a tried and tested programme;
* During an event there would be conflicting information;
* Various data sources were needed;
* The meteorological community had a stronger data-sharing ethos than geoscience at large;
* Lack of enforced data formats and a plethora of standards may pose a problem for real time operational effectiveness.

#### 2.14 Information management challenges associated with big data (T Logar (UN GlobalPulse))

1. T. Logar (UN GlobalPulse) outlined issues arising from the use of “big data”. He outlined several applications of big data and analytics, such as the use of mobile phone movements to deduce traffic flow. An experiment in Indonesia had used social media and citizen reporting to track haze events (a significant health problem in that region).
2. Typical challenges in the use of “big data” and artificial intelligence were security, privacy, ethics, “confirmation bias” (the human tendency to seek out views that conform to one’s own, or for a statistical correlation to be applied to make judgements on an individual), and technical issues.

#### 2.16 Requirements for support to the unified service interface (M Androli (WMO))

1. M. Androli (WMO) presented the concepts behind the Common Interface for Service Delivery (CISD) that would provide a one-stop-shop for service-related data products from different sources to meet new and evolving demands of users for real-time and intuitive access to weather and climate information as part of an integrated seamless service delivery.
2. The intention behind the CISD was to avoid the inconsistencies between the wide variety of weather information services available to users and the authoritative warnings from designated national authorities (the NMHS for weather warnings). By providing a standard interface to warning and other products, NMHS would make it possible for the large-scale data publishers to use the NMHS information.
3. The CISD was expected to:

* Serve to enable NMHSs to easily publish data through web services using technology compatible with mobile phones;
* Provide an interface to interrogate different forms of data from NMHSs;
* Enable user to access gridded data;
* Provide improved quality data.

### 3 Improving access to reference climate datasets

#### 3.1 Identifying reference datasets for climate (A Klein Tank (Netherlands))

1. A. Klein Tank (Netherlands) presented the Climate Services Information System (CSIS) climate services toolkit that supported development of climate services at national, regional and global levels. It had been recommended by technical experts from the Commission for Climatology, and was be available in the public domain at no cost to users.
2. The climate services toolkit (<http://www.wmo.int/cst>) included an inventory of GFCS relevant data and products that were available from WMO global and regional centres and other major climate institutions and provided links to many sources of information.

#### 3.2 The Copernicus Climate Change Service Global Land and Marine Observations Database: Plans, progress to date, and implications (P Thorne (Ireland))

1. P. Thorne (Ireland) described the Copernicus climate change service global land and marine observations database. The scope of the database was: to provide global land and marine surface meteorological holdings of multiple meteorological parameters across a range of time scales (sub-daily to monthly), with the intellectual property rights associated with data clearly articulated and flagged, and that were provided under a common data model.
2. The intention of the global land and marine observations database was to complement, rather than replace, national and regional data holdings, allowing access to global, historically collected, data in a single format. This would make it easier to calculate indicies and indicators that were built on multiple ECVs or to undertake analyses that required multiple variables. It would also enable sustainable data stewardship for a fundamental climate data record.
3. In the past, efforts to collect had narrower scope, so there were grossly duplicated holdings using a range of formats and metadata protocols. The challenge was to unify these into a coherent and consistent global data holding across all ECVs and timescales.
4. ICOADS was the recognized international repository for surface marine observations, with more than 30 years of heritage, data from ships, Buoys, fixed platforms and other sources. Data had been gathered from many sources, and eliminating duplicates was not always successful. In addition, some aspects of quality control were considered problematic. It was also dependent on a single institution for funding and staff.
5. The global land and marine observations database was being built by a team consisting of Maynooth University (Ireland), the UK Met Office, the UK National Oceanography Centre, the UK Science and Technology Facilities Council and NOAA’s National Centres for Environmental Information (USA).
6. The strategy for building the global land and marine observations database was for marine data to archive the ICOADS data sources and individual data elements, and for land data to source from a range of international, regional and nationally available sources – only public-facing sources were being used.
7. A detailed inventory of the available information was being prepared for publication on the service website (part of the broader climate.copernicus.eu domain).
8. A significant challenge was matching metadata to data, for which systematic platform identifiers were being used. The development team was discussing with WMO how WIGOS identifiers could be used.
9. Care had been taken to ensure that the data in the database could be shared and used by recipients. For data retrieved from ICOADS, only those that had no limitations on use were considered. For land stations the issue of data policy was more complex. The project was documenting the intellectual property rights status of all the data it used, and this would be used to control distribution of the data.
10. A common data model was under continual development so that all relevant metadata and data could be retained; this was intended to be compatible with WMO standards (and ISO 19115).
11. During data harmonization, quality controls were applied to the data, but no data were being deleted. This built on work by the National Oceanographic Centre and the Met Office and included; improved duplicate identification and flagging, improved handling of mis-positioned reports, improved tracking handling, better quality control and quality assurance and improved metadata. The project was also revisiting data sources and re-extracting data.
12. Techniques, computational capabilities and insights were developing all the time. This meant that reference datasets were unlikely to have a life beyond 30-50 years. Despite this, the underlying data from which they were created were irreplaceable. Therefore, and key responsibility was to recognize that the fundamental data record had to be retained permanently, and providing clear and responsible stewardship, withan audit trail of adjustments and quality decisions (while always retaining the original value so that it could be re-analysed at a later time).

#### 3.3 Copernicus/ C3S pulling together climate data, products and services (K Marsh (ECMWF))

1. K. Marsh (ECMWF) described the Copernicus climate change service (Copernicus/C3S). Copernicus was the EU’s earth observation programme, and was directed by EU and ESA (<http://www.copernicus.eu/>). It consisted of three components: space measurements, in-situ measurements and services to users. There were six services to users (each addressing a thematic area) that had been funded by the EU. ECMWF was the entrusted entity to run the Copernicus Climate Change Service (C3S) and the Copernicus Atmospheric Monitoring Service (CAMS). The Climate Data Store (CDS) would be at the heart of the C3S infrastructure and would provide information about past, present and future climate in terms of Essential Climate Variables and derived climate indicators, and many more variables, to increase the knowledge base to support adaptation and mitigation policies.
2. The CDS was being implemented to support a range of users (developers, subject experts, end users) through a unified web interface, using an operational framework that ensured data interoperability and standard services delivered by a distributed system.
3. Challenges faced by the CDS included: the variety of local and remote data stores with different access techniques, the data volumes (megabytes to petabytes), and the variety of data formats (netCDF, asci and shapefiles among others).
4. Development of the CDS used the agile methodology, in which there was an incremental and flexible approach with continuous stakeholder involvement. This allowed it to respond to evolving requirements. CDS was deployed on a cloud infrastructure to allow flexibility. It used standards-based components such as Geonetwork metadata catalogue that was compliant with ISO 19115 using INSPIRE rules and CSW (catalogue services for the web). The catalogue could be harvested using the OAI-PMH protocol that was used by the GISCs to manage the WIS metadata catalogue.
5. One of the major issues facing the CDS was “data granularity” – describing datasets in a generic way (“low granularity”) would return fewer records for a user to consider- but the descriptions of the data would be generic, making it difficult to select the most appropriate data. Highly detailed and specific metadata records (“high granularity”) would allow users to make a more detailed assessment of the datasets they discovered – at the cost of having to search through a far greater number of records. CDS requires data producers to determine an appropriate level of granularity for the metadata records describing their data.
6. The CDS had standardized contributions to C3S seasonal forecast data. Five institutions (ECMWF, UK Met Office, Météo France, Deutscher Wetterdienst and the Centro Euro-Meteterraneo sui Cambiamenti Climatici) were contributing seasonal forecasts from which products would be generated. In order to create a consistent archive, contributions were provided in netCDF 4 “classic” format using existing conventions (CF-1.6, SPECS, SMIP). The use of this format was standardized by adding additional definitions of: the exact structure for files, standard names for contents, standard spatial coordinates, multiple time coordinates and their use, realizations, cell methods, attributes and metadata. These were supported by a directory and file naming convention. The netCDF files could be checked for conformance to the CF-1.6 and C3S-0.1 conventions using a file checking tool at [ECMWF](https://software.ecmwf.int/stash/projects/CDS/repos/checker/browse).
7. A CDS toolbox was under development that would provide access to all the datasets available in the CDS and provide a simple way of manipulating them through web-based processing applications that used standard operations such as differencing, regridding, or statistical computations. These tools could be combined in workflows to generate end results for users.
8. Recommendations were:

* A standards-based approach for data and infrastructure was essential to achieve goals and maximise interoperability;
* If possible, contribute to the development and extension existing standards (e.g. CF);
* Close liaison with data providers, contractors, user communities etc. at an early stage was beneficial;
* An “agile” approach could be used for developing infrastructure.

#### 3.4 User interfaces for accessing information (J Tandy (WMO))

1. J. Tandy (WMO) discussed seven factors that should be taken into account when designing user interfaces: clarity on who the intended users were and on what they were trying to achieve; applying the principle that simple is best; it should not be necessary to provide a manual (but one should be produced anyway); do not assume that your user is an expert; clear navigation tools so that users can find their way around; in a multi-lingual environment provide the interface in English in addition to any other languages.
2. In discussion further points were raised: “put yourself in the shoes of the user”; people use computers and mobile devices in different ways, so the interfaces for these should also differ; provide machine access to the data to allow the data to be ingested directly to applications (this could be a bulk download service); The OGC/WWW Spatial Data on the Web Best Practices, (Open Geospatial Consortium and the World Wide Web Consortium, 2017) recommended that restful interfaces should be used in preference to Web Coverage Services, Web Feature Services and the like.

#### 3.5 Hydrohub – providing access to dispersed hydrological information (S Pecora (Italy))

1. S Pecora (Italy) the vice-President of the Commission for Hydrology outlined the Hydrohub that provided access to hydrological information.
2. The scope of the WMO Hydrological Observing System (WHOS) was defined as a multi-scale (local, national, regional and global) registry of hydrological data and information services catalogued using the standards and procedures of the Open Geospatial Consortium and the World Meteorological Organization.
3. While other interoperability studies had focused on implementing custom data streams (e.g. bridges, adaptors, etc.) between clients and server interfaces, WHOS focused on a common data and metadata management model that leveraged a suite of WMO and OGC standard Web services that could be applied to multiple scientiﬁc communities – in particular hydrologic and atmospheric sciences.
4. WHOS could be integrated within existing data discovery frameworks (e.g. portals, gateways, etc.) by leveraging mediation and brokering services.
5. Data publishers could register their data on the registry and provide brief descriptions of the datasets they wanted to share. This as an important aspect of WHOS because it allowed for data to be organized and discovered in an efficient, structured and methodical process.
6. In hydrology, the large number of activities and different applications had led to a tremendous heterogeneity of resources and procedures, which hindered cooperation among the different actors and stakeholders. This problem could be tackled through a mediated approach, that is identifying the existing heterogeneity boundaries and implementing suitable adaptation logic by means of specialized, lightweight components. More generally, the mediated approach relied on the identiﬁcation of articulation points around a particular heterogeneity boundary and the implementation of adaptation logic, whose execution is delegated to a specialized, lightweight component: the mediator.
7. The traditional approach for connecting diverse information systems based on commonly agreed standards did not scale up when addressing wide-ranging data interoperability. Interconnecting existing data systems had traditionally introduced limitations to their autonomy and scope, because different data providers may have different approaches to data and modelling and different vocabularies and even different interface protocols, bridging across providers was a more complex challenge. The brokering approach was formulated to handle such differences without limiting the autonomy and without putting a signiﬁcant investment burden on existing data systems. In particular, the brokering approach introduces a new middleware layer of service offerings.
8. The WMO Hydrological Ontology was a formal naming and definition of the types, properties, and interrelationships of entities that really or fundamentally existed in the domain of hydrology; in particular, it compartmentalized the variables needed in hydrology and established the relationships between them. The adopted development approach took into account not only semantics, but a more general perspective in order to address possible future needs. The conceptual challenge concerned the enablement of new ways of searching and the adopted methodology was essentially supported by the approach of enriching the searchable information that was associated with hydrological data and information.
9. The WHOS architecture as implemented by the combination of its components as a services stack framework providing a system in which data consumers could readily discover and access hydrological data, including time series, using spatial, temporal and semantic ﬁlters. At the core of the services stack framework lay the metadata services which acted as middleware between the catalogue services and data services. Data services ultimately provided the user with the data they were searching for, whereas catalogue services allowed users to perform federated searches across multiple data providers. Metadata services linked both these layers together by being registered at the catalogue level and providing all the information needed to access information at the data level.
10. The architecture of WHOS was also designed to offer advanced functionalities to all the global data centres, facilitating collaboration between NHSs and the centres themselves. A common benefit was the improvement of data discovery and access for both hydrological observations and products, available from Member States and the global centres as well. Furthermore, the published interfaces of WHOS offered a valuable technological evolution to global centres, fostering their operational activities in support of the Commission for Hydrology.
11. Experience in developing the WHOS was that when sharing information across scientiﬁc communities, it was important to deﬁne a standard framework through which large quantities of multidisciplinary information could be shared, discovered and accessed.

### 4 Plan for improving access to climate datasets

1. Break-out groups considered two topics: how to select reference climate datasets and how to promote the reference climate datasets.

#### 4.1 Selecting reference climate datasets.

1. Different applications had different requirements for data to support them. It was unrealistic to expect a list of reference datasets to be universally applicable. Further, many users would need to use multiple datasets to explore the uncertainties that would impact on their applications. The concept of a reference climate dataset was therefore thought to be too restrictive. Selecting specific datasets for inclusion in a list of “reference datasets” would not only be a never-ending process, but would also require some form of selection committee to decide on which datasets should be included in the list.
2. Instead of reference datasets, the breakout groups proposed the concept of a “trusted dataset”. The mechanism for expressing the “level of trust” for a dataset would need to be defined, but would be likely to include: the geographic and temporal coverage, the maturity of management of the dataset (see, for example, (Peng, et al., 2015)), the format of the dataset (digital was preferred), metadata describing the dataset, documentation of the data and their use, usage restrictions (the fewer the better), discoverability, how versioning was applied to the dataset, quality control and assurance, uncertainty estimates peer review of the dataset, and scientific credibility of the producers. User feedback mechanisms would allow “crowd quality control” that could increase the quality of datasets. A suitably defined set of metrics to express the level of trust for a dataset might allow a self-calibrating system to be implemented to support users in discovering the datasets most relevant to them.
3. The initial recommendation from this topic was that a method of defining the level of trust in a dataset would meet the objective of defining a set of reference climate datasets.

#### 4.2 Promoting reference climate datasets.

1. Discussions on how to promote reference climate datasets took place in parallel with those on defining which datasets should be identified as reference datasets. To achieve consistency with the conclusions on selecting reference climate datasets, this report of the conclusions of the breakout groups on promoting the reference datasets uses the phrase “most trusted datasets” wherever the breakout group used “reference datasets.”
2. The key tool to promoting the most trusted datasets would be including sufficient information in the metadata records describing to allow search algorithms to promote the datasets in their results. With such information it would then be possible to use searches of standard catalogues (such as the WIS catalogue) to prepare listings of the most trusted datasets for application areas. It was recommended that those listings should be linked from the home pages of GISCs (Global Information System Centres). Although based on the ISO 19115 geographic metadata standard, the information in the listings should be presented in a human-friendly way. In addition to listing the trusted datasets in searches of specialist catalogues, the datasets should also appear in the results of searches in the general purpose search engines.
3. The criteria for a high level of trust should include the use of open, documented formats for the data, and the quality of documentation of the dataset (including documentation of how to access the data). Both these aspects would be essential for the credibility of the ranking system; without such credibility the work expended in promoting the datasets would be wasted.

### 5 Identify existing information management practices

#### 5.1 Guidance for managing large volumes of information (J Pèrés (France))

1. J. Pèrés (France) discussed ways of managing large volumes of information, using the experience of Météo France as an example.
2. Météo Frace received 300 million observations a day, equivalent to 300 GBytes. The observation archive occupied 19 TBytes; the retention period depended on the type of observation with a minimum retention period of three months and some observations retained permanently. The climate dataset held minute, hourly, daily, monthly and norm data for meteorological and hydrological stations and associated metadata, a total volume of 700 GBytes. The operational forecast system, with a nest of three numerical weather prediction models, an additional limited area model for French territories outside Europe, an ensemble forecast system with a global (35 member) ensemble and national (12 member) ensemble, ocean wave models, an air pollution transport model and a seasonal forecast run once a month; together these generated on average 900GBytes a day with a total store of 5TBytes. The databases storing this information were built on PostgreSQL.
3. With data volumes increasing, there were plans to moving to a central data archive based on the ECMWF MARS system in which the data could be controlled more easily and storage used more effectively. The total archive size was 100PBytes; this was expected to exceed 250PBytes within two years.

#### 5.2 Guidance for managing international precipitation data (M Ziese (Germany))

1. M. Ziese (Germany) described how global precipitation data were managed by the Global Precipitation Climatology Centre (GPCC). The GPCC had been responsible since 1989 for global collection and analysis of in-situ data for land-surface precipitation. It contributed to the Global Energy and Water Exchanges project and to GCOS. Its sources of data includd SYNOP and CLIMAT reports from the NOAA Climate Prediction Center, the European Climate Assessment and Database, the University of East Anglia Climate Research Unit, the Food and Agriculture Organization, the Global Historical Climatology Network, national meteorological services and regional data collections.
2. Data were received in many different formats. Quality control was applied to these and they were transformed into a standard format within a database from which they could be extracted for further analysis.
3. There were many possible ways to structure the data in a database, the main items that could be used to index that structure being provider, parameter and station. Taking into account the opportunities and disadvantages of each of the storage structures, the GPCC stored its data using the hierarchy: station, provider, parameter, date.
4. Quality control was a major component of data management. The original data were retained, and all quality control decisions were recorded in the database. Care needed to be taken when interpreting quality control information – a value that may appear to be in error may actually be a description of an extreme event. Both manual and automated quality control methods were used. Quality control led to one of three outcomes: the data were confirmed (flagged values corresponded to extreme events), the data were corrected (the original value was incorrect and was replaced in the “current” data by a more realistic value – an example would be if a reported value had a misplaced decimal point); the data were deleted from the “current” set (the original value was obviously wrong but there was no information that would allow it to be corrected).
5. Station metadata were stored as a station history with each element being recorded with the period for which it was valid. The approach taken by GPCC meant that the location of a station could change during a time series.
6. For precipitation data, it had been found that the most effective methods of managing data were to:

* Store data in a station-specific data bank;
* Store station history and related meta-information alongside the observations;
* Store data separated according to provider to allow for cross-comparison;
* Apply advanced quality control procedures;
* Allow for roll-back of any QC related changes in the data bank, and keep a copy of the original data
* Use the best available data at each time of an analysis;
* Respect and document copyright and intellectual property rights of data providers

#### 5.3 Guidance for managing climate data and scenarios (R Dunn (UK))

1. R. Dunn (UK) described the approach used by the UK Met Office to manage national and global climate observations and simulations.
2. The Met Office was the data provider for UK climate data. Two key principles were followed when preparing climate reports: traceability and reproducibility. Climate records needed to be linked through the processing chain to the original report – that could be in manuscript. The production of UK statistics was through a portable, modular and traceable software system built to open software standards.
3. The value of climate datasets and studies lies in their application. Easy and rapid access to reference datasets was fundamental to climate monitoring to support national and international climate assessments. Not only was it essential to have access to other sources of data when preparing reports, but it was also necessary to publicize as widely as possible the availability of reports generated by the centre.
4. The Met Office Hadley Centre maintained over 15 climate datasets, with further datasets archived because they had been used to support earlier studies. Each dataset had a nominated point of contact. The datasets were available through the Met Office web site.
5. Versioning of datasets was important to support reproducibility of results and identification and correction of errors. The Met Office was working to a three-level version numbering system in which the most significant level was incremented when a dataset and its production process had been peer reviewed, the second level when there has been additional internal review and documentation and the third level when the contents had been expanded (for example by adding an additional year of data).
6. Quality control of climate information was a potential source of error and controversy. The Met Office was working towards an open approach to quality control, with the procedures and software made publicly visible through Github. The aim was for open code review during development of a dataset, with the code under version control and the inputs and outputs of each dataset creation activity being stored. The software itself was intended to be self-documenting, reducing the likelihood of discrepancies between the code and its documentation. Developing the code in public allowed for wider review (and thus more errors being identified before the dataset was created).
7. Consistency in analysis and creation of datasets was important. Code was written using open standards and tools that were themselves made open. The production process itself was automated, making it intrinsically more repeatable.
8. The Met Office Hadley centre stored its climate data in asci and netCDF formats, but even storing data in compliance with standards still left scope for variations. Cooperating with other centres increased the variety in formats – production of the WMO annual statement involved five data centres – and far more than five data formats.
9. Controlled vocabularies were needed to make sure that data from different centres actually represented the same thing. Using a common vocabulary simplifies the task of transferring data between formats.
10. The Met Office Hadley Centre was working on further improving its processes. Open processing relied on visibility of the data, which was not always possible if the data used were subject to intellectual property rights (IPR) restrictions. Clarity of the IPR constraints (licensing) associated with data was important. Identification of datasets, through Digital Object Identifiers, especially for datasets that were updated frequently, would assist with traceability. In the past, station metadata had been stored separately from the data, making it difficult to associate the two; this was being addressed. The quality assurance procedures were also being enhanced.
11. Climate scenario information was a combination of observational data (gridded, point, track or profile) that was often incomplete, and model data that were usually gridded and always complete. Each source had associated uncertainties. The latest scenarios produced (UKCIP18) were provided in a number of forms depending on the intended application; some consisted of data only, while others contained selected data and associated guidance.
12. Recommendations for managing climate data were:

* Endure traceability – from raw observation (digital or manuscript) through the climate product;
* Automate for improved reproducibility;
* Provide open and easy access to climate data to ensure rapid and clear communication of information, results and analyses;
* Use a clearly defined versioning scheme to allow exact identification of dataset used, tracking of updates, and correction of errors;
* Use open code under version control to allow knowledge exchange, documentation, code review and collaboration;
* Provide documentation for clarity and consistency of approach;
* Peer review of the methods used to process data is vital in ensuring the scientific integrity of the work;
* Use external collaboration and hosting to improve resilience to external influences;
* Use standard formats for data to simplify data conversion (rather than introducing more standards);
* Use a controlled vocabulary with the standard data formats;
* Tailor outputs for different users where appropriate, allowing for basic to advanced access methods.

#### 5.4 Guidance for managing marine information (E Freeman (USA))

1. E. Freeman (USA) described guidance for managing marine information. This was based on the WMO-IOC strategy for data management.
2. The JCOMM climate data management system was not standardized, complex, poorly documented, not well linked and poorly understood. The vision for a modernized marine climate data system (MDCS) was to: modernize the Marine Climatological Summaries Scheme MCSS; standardize international data management across JCOMM; integrate marine-meteorological and oceanographic data; store both real-time and delayed-mode data and the supporting metadata; incorporate new sources of historical data; and use state-of-the-art data management.
3. The MCDS would operate through three tiers of centre: data acquisition centres (DACs) would receive the data for a single platform type in real-time and delayed modes and create standardized output with only basic quality control; global data assembly centres (GACs) would process data from one or more platform types, combining data and metadata from the DACs, homogenize the data formats and perform higher level quality control; and centres for marine-meteorological and oceanographic climate data (CMOCs) would handle data from many platform types, combining their data and metadata into a common format, providing international mirroring of data holdings, leading data and metadata rescue activities, provide value-adding processing (for example adjusting for biases), produce climate products, and maintain a long-term archive.
4. Recommendations for managing marine information were:

* Implement better connections between TT- / ET-MOWIS and the Expert Team on Marine Climatology (ETMC) and TT- Marine Climate Data System (TT-MCDS) on connectivity and data discovery;
* Maintain MCDS representation on the TT/ET-MOWIS team to facilitate information management between the WIS and MCDS;
* Promote development of the Marine Climate Data System (MCDS) regarding the Joint WMO and IOC Strategy for Marine Meteorological and Oceanographic Data Management (2018-2021) and assist in promoting the MCDS and making data discoverable;
* Support CMOC development and connections to the WIS;
* In collaboration with ODIS/ODP and WIS and JCOMM ETMC, update marine meteorological and oceanographic publications with procedures for connecting to portals and discovery services as systems develop;
* Clearly define how WIS 2.0 interacts with the Ocean Data Portal and how portals interact.

#### 5.5 Guidance for managing national climate observations (B Bannerman (Australia))

1. B. Bannerman (Australia) outlined the Climate Data Management System Specification that had been provided as guidance (World Meteorological Organization, 2014). Although this provided good guidance, further development was needed to support developers of climate data management systems (CDMS) to reflect additional requirements for climate information. In addition, additional work was needed to develop agreed data policies, formal data definitions and other infrastructure to allow a federated global climate data services to be delivered.
2. Custodians of climate data faced a challenge: “An important point to remember is that the most important stakeholders for our data have not been born yet and that we are custodians of this data for their future use” *(B. Trewin).*
3. Climate data were being used by many organizations to support a wide variety of applications, including disaster resilience, energy and mineral resource management, infrastructure and transport management, biodiversity and ecosystem sustainability, and sustainable urban development.
4. In 2012, a survey of countries on CDMS had shown that 48% of respondents had operational issues (and that 10% of systems were not operational), 25% did not use a database, 19% relied on spreadsheets to manage climate data, 45% had developed their own CDMS, and 50% would like to change their systems. There were 96 different climate data management systems in use in the 137 countries responding to the survey.
5. Although there are 96 different CDMS in use, they provide different facilities and probably none of them complied fully with the specifications. Forty-two countries relied on an open source climate data management system for which there were limited funds and skills available for development and maintenance.
6. The Commission for Climatology’s ET-CDMS had recommended that open source CDMS should converge on one reference open source CDMS.

#### 5.6 The Integrated Meteorological Information Service System in CMA: Progress and Outlook (F Gao (China))

1. F. Gao (China) described the approach being taken by CMA to providing a centrally managed repository for meteorological information.
2. The data were stored in CIMISS( China Integrated Meteorological Information Service System). This was used to collect data from the whole of China and to provide a high-quality data service for climate. The experience in China was that from the implementation of data management, to accommodate requirements of climate services in information management system, the following points should be addressed:

* The primary design principle of the database system should be service oriented;
* A uniform framework and attributes of the database systems at different levels a the national and local level simplify development of applications and using them in combination;
* The real-time and historical data had both to be included within an integrated database, to meet climate service requirements;
* Using a persistent unique identifier for each item of data, and maintaining a uniform data classification and coding system within the in database system, permitted traceability of data throughout its lifecycle, from observation to data service;
* A single unified service interface to data (e.g. CMA’s MUSIC) was necessary to provide high-quality data service, supporting multiple applications, operating platforms, development languages, and programming habits;
* To intend to work on collecting data from different sources, we have to focus three points, data normalization, data automation processing and data checking.

#### 5.7 Use of a maturity matrix to guide management of multi-disciplinary information (C Lief (USA) (remote presentation))

1. C. Lief (USA) described the use of the maturity model for data stewardship to guide management of information at the NCEI, (Peng, et al., 2015).
2. The data stewardship maturity matrix had been developed to measure stewardship practices that were applied to individual digital environmental data products that were publicly available online. It was developed to help NCEI make sure that it was complying with national directives for public environmental data, and addressed ensuring that the data were preserved and secure, available, discoverable and accessible, credible and understandable, usable and useful, sustainable and extendable, and citable and traceable.
3. The matrix covered nine components that could each be assessed against a five level maturity measure. The components covered were preservability, accessibility, usability, production sustainability, data quality assurance, data quality control/monitoring, data quality assessment, transparency/traceability, and data integrity. Each was assessed against the levels: 1 – not managed, 2 – managed limited, 3 – managed defined, partly implemented, 4 – managed well-defined, fully implemented, level 4+ measured, controlled, audit. Each component/maturity combination had descriptions of behaviour that were expected to be achieved. The model was described in more detail in (NCEI, 2015).

#### 5.8 Guidance for managing crowd sourced information (H Reges (USA))

1. H. Reges (USA) described how crowd sourcing rainfall information by providing simple rain gauges to volunteers had created over 20 000 observation sites in the USA.
2. Five areas had been identified to describe best practice in the management of sources of information from volunteers: deciding on what data you would like to capture; capturing and displaying the data; quality controlling the data; ingesting and storing the data; disseminating and sharing the data.

#### 5.9 Guidance for managing international cryosphere information (D Gallaher (USA))

1. D. Gallaher (USA) described the experiences in managing information about the cryosphere.
2. The practices discussed during the talk were: standards and formats, data quality, data accession and de-accession, branding, data set versioning guidelines, data set DOI guidelines, data set citation guidelines, issues and limitations, legal concerns, example cryosphere data, and issues.
3. Standards and formats used by the cryosphere community included:

* Representation: GeoTiff, HDF, NetCDF, binary shape and KML;
* Access protocols: OpenDAP, FTP, HTTP, HTTPS, OGC Web Services;
* Metadata: ISO 19115;
* Catalogue services: OpenSearch, OAI-PMH.

1. Data quality was assured using the QA4EO principles of quality indicators, traceability, the reference standard used and uncertainty.
2. Data branding was important to help users find the data they needed.
3. Guidance for the cryosphere community on DOIs was to assign a DOI when a dataset was first published (even if it was under embargo on publication) and to issue a new DOI for each new major version (a version in which the scientific content changed). Both DOIs and the metadata describing a dataset were retained even after the dataset itself may have been destroyed.
4. Key issues in information management were: factors beyond metadata and standards must be considered in dataset selection; there were factors beyond costs that could block open data access; Sea Ice index and IceSat II could be considered as data sources for general use; when deciding on reference datasets it was necessary to understand the intended audience for them.

#### 5.11 ICAO System-Wide Information Management (A Moosakhanian (USA))

1. A. Moosakhanian (USA) described the practices and plans for managing aviation information in the Federal Aviation Administration and the related activities within ICAO.
2. He outlined the concept of the System-Wide Information management System that would allow information from the different aspects of aviation to be combined using industry-standard techniques to develop more-powerful end user applications. In the USA this was supported by the Common Support Services (Weather) and the NextGen Weather Processor.

## Annex 4: Bibliography

# Bibliography

Intergovernmental Oceanographic Commission, 2017. *IOC Strategic Plan for Oceanographic Data and Information Management(2017-2021).* Paris: Intergovernmental Oceanographic Commission.

International Civil Aviation Organization, 2014. *Manual on the Digital Exchange of Aeronautical Meteorological Information.* Montreal: International Civil Aviation Organization.

International Civil Aviation Organization, 2015. *Meteorological Service for International Air Navigation. Annex 3 to the Convention on International Air Navigation..* Nineteenth ed. Montreal: International Civil Aviation Organization.

International Civil Aviation Organization, 2016. *Global Air Navigation Plan.* 5 ed. Montreal: International Civil Aviation Organization.

Mosely, P., 2001. *Exchange of hydrological data and products, technical reports in hydrology and water resources.* Geneva: World Meteorological Organization.

NCEI, 2015. *The scientific data stewardship maturity assessment model template,* s.l.: NCEI.

Open Geospatial Consortium and the World Wide Web Consortium, 2017. *Spatial Data on the Web Best Practices.* s.l.:World Wide Web Consortium (W3C).

Peng, G. et al., 2015. A unified framework for measuring stewardship practices applied to digital environmental datasets.. *Data Sciene Journal,* Volume 13, p. 23.

World Meteorological Organization and Intergovernmental Oceanographic Commission, 2017. *Draft Joint WMO and IOC Strategy for Marine Meteorological and Oceanographic Data Management (2018-2021).* s.l.:World Meteorological Organization.

World Meteorological Organization and Intergovernmental Oceanographic Commission, 2017. *The Ocean Data Standards and Best Practices Project (ODSBP).* Paris: World Meteorological Organization and Intergovernmental Oceanographic Commission.

World Meteorological Organization, Japan Meteorological Agency, Deutscher Wetterdienst, 2012. *Assessment of the observed extreme conditions during late boreal winter 2011/2012.* Geneva: s.n.

World Meteorological Organization, 1994. *Data acquisiiton and processing, analysis, forecasting and other applications. Guide to hydrological practices..* 5 ed. Geneva: World Meteorological Organization.

World Meteorological Organization, 2010. *Assessment of the observed extreme conditions during the 2009/2010 boreal winter.* Geneva: World Meteorological Organization.

World Meteorological Organization, 2013. *The Global Climate 2001-2010: a decade of climate extremes.* Geneva: World Meteorological Organization.

World Meteorological Organization, 2014. *Climate Data Management System Specifications....* s.l.:World Meteorological Organization.

World Meteorological Organization, 2016. *The Global Climate in 2011-2015.* Geneva: World Meteorological Organization.

World Meteorological Organization, 2016. *WMO Statement on the state of the global climate in 2016.* Geneva: World Meteorological Organization.

World Meteorological Organization, 2016. *WMO Techical Regulations, Basic Documents No. 2, Volume II - Meteorological Service for International Air Navigation..* 2016 ed. s.l.:World Meteorological Organization.

World Meteorological Organization, 2017. *WMO Information System (WIS) 2.0 strategy.* s.l.:World Meteorological Organization.

--------------------

## Annex 5: Topics for which guidance on information management would be beneficial

The following table lists the topics that the workshop breakout groups identified as needing guidance, and the resources that could be used to support development of the guidance.

| **Topic** | **Resources to support development of guidance** |
| --- | --- |
| Metadata | NOAA guidance  Rubric  Other USA bodies  WIS Guide part ?  WIGOS metadata standard  model metadata – netCDF guidance  Services metadata?  Catalogue level metadata, parameter metadata  CDMS specifications (also covers other topics) |
| Metadata management | Guides are available |
| Metadata format | Evolve format as go forward – interoperability with “other” repositories  Cross-walks |
| Metadata quality |  |
| Provenance | NOAA metadata guidance |
| Versioning (recording in metadata) |  |
| Versioning (of metadata) |  |
| Traceability | QA for EO standards: http://qa4eo.org/ |
| System interoperability | Manual on WIS / ODP / SWIM |
| Curation of related artefacts (documentation, algorithms, maps…) | ? UK National archives  ISO 19165 and its references |
| Intellectual Property Rights |  |
| Licensing | Res 40, 25, 60, CC, open data … Need summary guidance |
| Data policy | Res 40, 25, 60, open data |
| Open data | INSPIRE guidance |
| Closed data | WIGOS TT-WDP should provide guidance  CBS-LR-EDI recommendations |
| Recording alongside data/metadata |  |
| Respecting IPR in use/distribution of data | “Need systems in place to respect IPR of others”  “Make sure you record and inform when you change licence for your data” |
| Tracking IPR | WIPO guidance |
| Security | WMO guide on ICT security |
| Authentication/authorization |  |
| Permissions management |  |
| Integrity |  |
| Availability |  |
| Backup/recovery | WWW guide to data management (principles) |
| Mirroring of data/metadata holdings |  |
| Preservation | CCl HQDMFC CONOPS (covers other topics too) |
| Long term availability of services |  |
| Integrity |  |
| Archive |  |
| Data rescue |  |
| Quality |  |
| Quality control |  |
| Feedback mechanism to improve quality |  |
| Data |  |
| Versioning |  |
| data format | has to be stated and recognized and self-describing Manual on Codes, netCDF … (managed diversity rather than chaotic diversity) |
| interoperability |  |
| data management | Manual/Guides to marine met services  CONOPS for HQDMS, WWW data management guide, WIGOS management of partner data, |
| Quality |  |
| Crowd source data | CBS-LR-EDI identify issues to be taken into account (need to define a scope) |
| Information from partners |  |
| Unstructured data |  |
| Model vs observation |  |
| Retention period | NCEI guidance, UKMO guidance, BoM guidance, MF guidance – thought processes when setting retention periods  Primary observational data **shall** be retained permanently (for climate). |
| Citation |  |
| Processing | W3C guidance on interpolation, sub-setting |
| SGDPFS |  |
| Quality control need statement on continuous improvement based on feedback on assessments of data (issue guidance from Commissions) | Ocean data standards and best practices, WIGOS Manual/Guide, CIMO guide, CDMS specification, Manual on GDPFS (data assimilation too?), several climate-related quality management, HQDMS manual. WCRP CMIP process, GAW statements,  DWD software for marine data and associated manual. OPERA guidance  Manual on GDPFS – verification / inter-comparison of models. |
| Homogenization |  |
| Applications and tools |  |
| Management processes | UK Information management model, BoM Framework on information management, |
| Maturity matrix | NCEI / CoreClimax (data products) / measurement system (GAIA-clim) |
| Search and discovery |  |
| Search protocols |  |
| Intended audience |  |
| user can assess suitability for purpose  human access  Machine access | Copernicus EQC contracts will address from climate perspective – extensible?  INSPIRE addresses  W3C management of data (3-4 best practices) |
| usability index |  |
| controlled vocabularies (crosswalks) | GCMD, NetCDF\*n, Meteoterm, BUFR tables, codes.wmo.int, |
| User interface/user experience |  |
| “trustability” |  |
| Sharing and access | Spatial data on the web best practices (and covers other of our headings)  Manual on WIS/GTS, JCOMM… |
| Designation of authoritative sources (this includes authoritative aggregators as well as the data originators) | Manual on GDPFS, Manual on WIS, JCOMM guidance on WIS centres, OSCAR/surface, Manual on WIGOS, ICAO regulations, |
| Real time |  |
| Delayed mode |  |
| System interoperability |  |
| Where to share data | “Market place” for data.  Need to make it clear where the original data are to be found – authoritative source (synchronize copies with authoritative version – including deprecation). |
| How to share your data | guidance on sharing national data with academic and other areas, and also with international community (some application area, some WIS, some policy guidance)  WDCs should document how to get data to them. standard approaches? |
| User interface/user experience |  |
| Acquisition | Manual GTS, WIGOS Manual, Manual GOS, CIMO guide, Manual GDPFS?, “Hydrology guide” current guidance is inadequate (people say acquisition is too difficult)  Manuals and Guides on Marine Met Services (plus other topics)  GRUAN manual and guide & tool for collection of “deep metadata” RS Launch client. |
| Check acquired data meet requirements |  |
| Initial provision to publishing centre |  |
| Sending data to designated “world data centres” |  |
| Data creation |  |
| ICT infrastructure |  |
| Technology |  |
| Financial management |  |
| Services in the cloud |  |
| Cloud solutions |  |
| Availability management |  |
| Change management |  |
| Capacity management |  |
| Managing hardware | WIS competences |
| Managing environment | WIS competences |
| Managing operating systems | WIS competences |
| Managing applications | WIS competences |
| Managing storage | WIS competences |
| Managing telecoms | Manual GTS/WIS and guidance |
| Managing outsourced services |  |
| Managing … |  |
| Training |  |
| Documentation |  |
| Training |  |
| FAQ/help |  |