WORLD METEOROLOGICAL ORGANISATION

Climate Data Management System (CDMS) Specifications

WMO-No. 1131

Edition 2024

[1] http://creativecommons.org/licenses/by-sa/3.0/

Publication revision track record

Date	Part / chapter / section	Purpose of amendment	Proposed by	Approved by

Table of Contents

Acknowledgements

The original version of this document (2014 edition) was produced under the leadership of Bruce Bannerman from the Bureau of Meteorology, Australia, and Denis Stuber from Météo-France, whose outstanding contribution to this publication, including its completion and finalization, is gratefully acknowledged.

The revision on WMO No 1131 was undertaken by Denis Stuber (Météo- France) and William Wright (retired; formerly Bureau of Meteorology, Australia), with the valuable assistance of Peer Hechler (WMO Secretariat).

The following people have also contributed to this publication: Bruce Bannerman (formerly Bureau of Meteorology, Australia), Gao Feng (China Meteorological Administration, China), Joël Guesquin (Météo-France, France), Jose Antonio Guijerro (formerly State Meteorological Agency, Spain), Dave Inglis Berry (WMO Secretariat), Jörg Klausen, (MeteoSwiss, Switzerland), Arun Kumar (Climate Prediction Center, USA), Ge Peng (University of Alabama in Huntsville, USA), Radim Tolaz (Czech Hydrometeorological Institute, Czech Republic), Alessandro Spinusso (Royal Netherlands Meteorological Institute, Nederlands), Blair Trewin (Bureau of Meteorologie, Australia).

1 Executive Summary

This publication represents an update of the Climate Data Management System Specifications originally published in 2014. It presents an updated framework defining the functionality required or recommended within a climate data management system (CDMS). A CDMS is an integrated computer-based system that facilitates the effective acquisition, archival, management, analysis, delivery and utilization of a wide range of integrated climate data.

The framework comprises a set of interrelated building blocks called components. Each component describes a specific functional requirement of a CDMS and contains, where appropriate, references to further information. Components are grouped according to broad functional classifications, such as Governance requirements, Time-series analysis, general data management, etc., presented as separate Chapters. A diagram at the start of each Chapter summarises the components, and where appropriate, differences from the previous (2014) version, in terms of either new requirements, or changes in classification (next paragraph).

Not all components are needed within every CDMS. Components are classified as Required (i.e. mandatory), Recommended (i.e. recommended best practice) or Optional (i.e. considered a more advanced functionality). Appendix 1 contains diagrams that summarize this classification. Appendix 2 contains a list of the required CDMS components.

It is important to note that a CDMS is not expected to contain all of its functionality in a single software package. The components have been designed to group similar functionalities together. In many cases, the functionality of components can be provided by existing off-the-shelf, open-source and proprietary software applications. However, some effort will be required to integrate these components together.

National Meteorological and Hydrological Services (NMHSs) may find this component classification useful as a guide to implementing CDMS functionality that they can afford to maintain over the long term in order to effectively manage and use climate data. It should be noted that Regional Climate Centres may be able to provide more advanced services to less developed countries in the future, along with emerging technological opportunities such as Cloud technologies.

It should be noted that some of the components may be under the responsibility of another body or entity other than an NMHS.

It is expected that the World Meteorological Organization (WMO) and NMHSs will take some time to make the necessary changes to their CDMS. Chapter 10 outlines an approach that organizations could take to make this change.

This framework can be thought of as a taxonomy defining a common set of terms for CDMS components. It is anticipated that the framework will underpin future work to compare the functionality available in competing and emerging CDMSs.

An important aspect of the Governance section in particular (Chapter 3), but applicable to other Chapters as well, is that many of the components apply, or should apply globally. In this edition, this includes such fundamental WMO Resolutions as WMO Resolution No 1 on Data Exchange, which replaces a number of previous Resolutions. In addition, emphasis is placed on conforming to community standards, both within WMO (such as the WIGOS Metadata Standard), and outside (ISO Standards and OGC, etc.). This is to ensure that CDMSs conform with WMO Technical Regulations, infrastructure requirements, and the necessary interoperability that provides efficient data ingest and exchange.

This publication is a living document and will be updated on a periodic basis to reflect advances in technologies, new or updated guidance materials, new WMO Resolutions, etc.

This CDMS Specifications is intended for CDMS procurement staff of NMHSs, climate data managers, systems integrators and CDMS developers, architects, vendors, etc.

Definitions

Authoritative Data Source	An Authoritative Data Source is the accepted point of truth for the data given by an organization or person having the legitimacy in a concerned domain.
The Climate Record	The Climate Record constitutes the “official climate data” repository for an NMHS. It is the observational record of meteorological events recorded by observers or instruments according to accepted observation standards, and includes: • electronic and paper records, as well as scanned images • climate observations and metadata • data derived from both in-situ and remotely sensed observations • WMO standard climate products • homogenised data • output generated by numerical models • other related data such as proxy data. The data comprising the record consists of the “as read” variables, along with values following routine quality control processes. The Climate Record should be managed as a permanent long term repository of data that is pertinent to Climate usage.
Cloud Computing	The use of technology, data and services, software, etc. on the internet rather than software and hardware that are installed on the computer(s) of the user. Hardware, applications and data no longer reside in the premises of the user, but in a “cloud” made up of numerous interconnected remote servers.

Standard	Standards in the context of this publication are a common set of established guidelines, specifications, rules or schemas, approved by a recognised body. Standards enable consistent, widely understood communication and adoption between different entities by providing a common framework for handling and interpreting information. Examples of standards that are defined in this publication include the ISO 19289:2015 on the siting classification for surface observing stations on land.
Third-party	Third party is any entity, e.g. private providers, government agencies, universities, associations or individuals (amateurs, volunteers, ...), etc., that observes or measured or generates climate data and/or climate services and does not have a direct relationship with, in the case of this document, a National Meteorological and Hydrological Service (NMHS) or with WMO.
Third-party data	Are data obtained from third-party providers. The data may or may not conform to NMHS and WMO expectations, but ideally these aspects should be well documented.

2 Background

This publication represents a follow-up to the 2014 edition of WMO No 1131, itself the result of numerous requests for guidance on what are the requirements for Member NMHSs to effectively manage and utilize climate data, and which specifications should be considered for CDMS implementation. It builds upon the first edition by including additional functionality, updating references, and reflecting changing priorities in data management.

The concept of a CDMS described in this document goes beyond that of a coredatabase, where data are ingested, managed, archived and accessed, and is linked to the functionality required to cover all requirements for data- and time-series management (Chapters 4 and 5), under a sound governance framework (Chapter 3). Chapters 6 through 9 describe the functionality required for data analysis, presentation, delivery, as well as core IT infrastructure.

Using an IT architectural approach the publication describes, on an abstract level, the various logical components that make up the integrated system required to support CDMSs. It further classifies components according to their relative priority: required (i.e. mandatory), recommended (i.e. best practice) or optional (i.e. considered a more advanced functionality).

This framework can be thought of as a taxonomy defining a common set of terms for CDMS components. It is anticipated that the framework will:

- provide a gap analysis to highlight to Members where attention needs to be given to improve existing CDMSs;
- provide a basis for assessing suitable CDMSs for Member NMHSs currently without a properly-functioning CDMS or who wish to upgrade; and
- eventually lead to a generic CDMS solution.

This Specification is a living document, to which more detailed definitions of requirements will be added as they become available.

The CDMS description contained in this publication is intended to cover the deployment of any CDMS, from those in developing nations to ones in organizations with large computing resources. It is intended to inform the data management requirements not only of the Member NMHSs, but other scientific entities with responsibility for managing climate data, ranging from universities to Regional and Global Data Centres.

It is important to clarify several things:

- Firstly, these specifications are designed to align with the requirements and standards of existing WMO publications, as specified in:
  - ✓ *Technical Regulations* (WMO-No. 49),
  - ✓ *Manual on the High Quality Global Data Management Framework for Climate* (WMO-No. 1238),
  - ✓ *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160),

- ✓ *Manual on the WMO Information System* (WMO-No. 1060),
- ✓ *Guide to Climatological Practices* (WMO-No. 100),
- ✓ and with other WMO Standard documents, such as the *WIGOS Metadata Standard* (WMO No 1192).

- Secondly, a CDMS utilized is not necessarily expected to contain all of the functionalities described in this Specification, although every effort should be made to ensure that the Mandatory functions at least are incorporated.
- Thirdly, it is particularly important to note that the functionality need not be, and should not be, incorporated in one monolithic software package. The components have been designed to group similar functionalities together. In many cases, the functionality of components can be provided by existing off-the-shelf, open-source and proprietary software applications, or via Cloud-based technologies, or via WMO infrastructure initiatives such as the WMO Information System (WIS) and/or OSCAR/Surface for the documentation of the observations made. Some effort, however, will be required to integrate these components together. Some components, such as the Governance components identified in Chapter 3, are not physical components at all, but still required to ensure efficient, secure and trustworthy data management.

The target audience for this document includes: Climate and CDMS procurement staff of NMHSs, climate data managers, systems integrators and CDMS developers, architects, vendors and so forth. It is recommended that these stakeholders:

- Treat this Specification as a description of the functionality that is either required or desirable for their CDMS.
- Assess the capability of their current systems against the components outlined below, as discussed in Chapter 10.
- Plan their acquisition, procurement and development activities, as part of their maintenance strategy over three to five years, to ensure that they implement functionalities that are consistent with:
  - ✓ At least the mandatory functionalities described in this document;
  - ✓ Their level of technical expertise and funding.
- Ensure that they continue to fund CDMS activities at a level that is sustainable for the long term.

“An important point to remember is that the most important stakeholders for climate data have not been born yet and we are implementing and managing CDMSs to protect the integrity of climate data for their future use.”[2]

A theme of these specifications is to ensure that the data are securely and transparently managed.

[2] Statement attributed to Mr Blair Trewin, Bureau of Meteorology, Australia

2.1 The climate system

To understand the climate system, one must understand a wide range of complex interactions and interrelationships occurring over time. Figure 2.1 is a simplified pictorial representation of many of the aspects that make up a climate system.

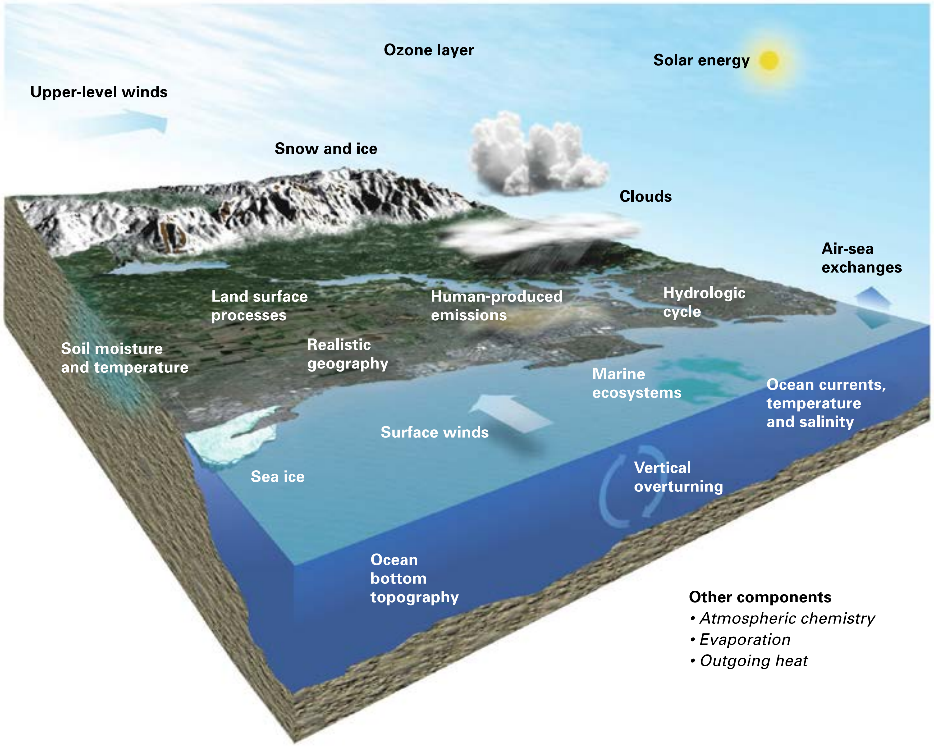


Figure 2.1. Community Earth System Model

Source: National Center for Atmospheric Research, United States of America. Image can be found at https://scied.ucar.edu/image/community-earth-system-model.

For a more detailed overview of the climate system, see:

- *Guide to Climatological Practices* (WMO-No. 100)

2.2 Observing operations for climate data

To understand what is happening in the climate system, a wide variety of observations of various climate-related phenomena are routinely made, stored, managed and analysed using CDMSs. Examples of instruments used to make such observations are portrayed in Figure 2.2.

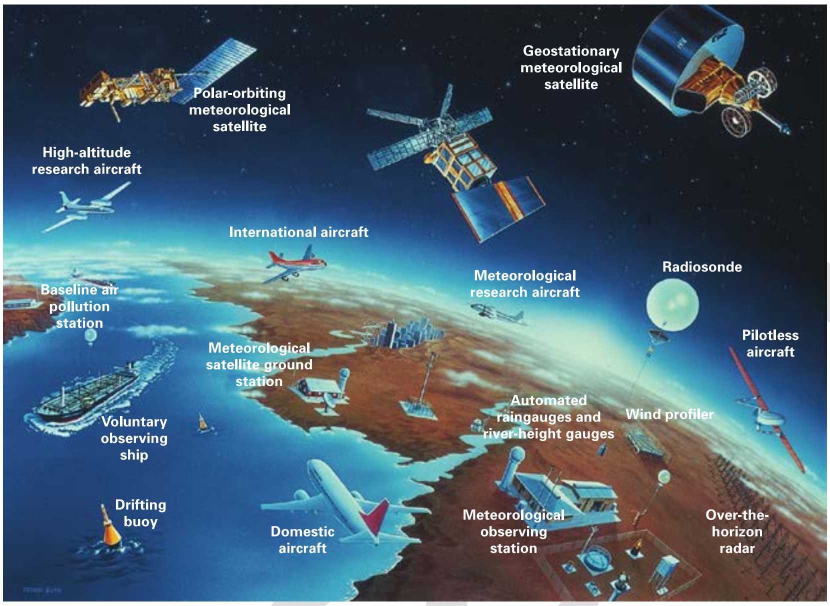


Figure 2.2. Multiple observing systems

Source: Adapted from a WMO flyer on the WMO Integrated Global Observing System (WIGOS)

These observations are routinely analysed and a range of derived data generated, for example via climate reanalyses. The significant amounts of data generated must be effectively managed to ensure that they can be easily and routinely used to help improve our understanding of the Earth's climate. This publication describes, on a conceptual level, an integrated framework that is necessary for effectively managing and using this climate data. Reference is made to enabling infrastructure for large and complex data systems, including the Cloud, and the capabilities of the WMO information System (WIS).

For a more detailed overview, see:

- *Guide to Climatological Practices* (WMO-No. 100), Chapter 2.
- *Guide to the WMO Integrated Global Observing System* (WMO-No. 1165)

2.3 Major components of a climate data management system

Figure 2.3 is a graphic depiction of the major functional components that comprise a CDMS.

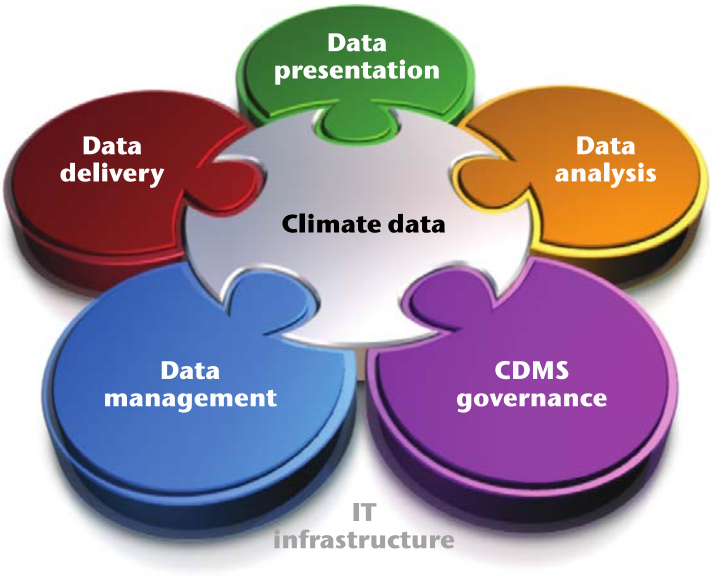


Figure 2.3. Major components of a climate data management system

A summary of these components is provided below.

2.3.1 Climate data

The Climate data component represents a wide range of time-series climate data. It extends well beyond what may be thought of as traditional meteorological observations and includes:

- The variables described within the Observing Systems Capability Analysis and Review Tool, the OSCAR variables database;
- The Global Climate Observing System (GCOS) Essential Climate Variables (ECVs), noting that these cover the domains of atmospheric, marine, hydrological, cryospheric and other domains;
- An expanded view of climate metadata that includes metadata on observations, discovery and data provenance, which latter is given more prominence in this publication than its predecessor;
- Standard WMO products;
- Derived observations and gridded data;
- Outputs from numerical models;
- A range of ancillary data used to support CDMSs, including spatial and impact (i.e. socio-economic effects of climate events) data, documentation and climate software;

- Other important ICT (Information and Communications Technology) underpinnings such as logical data models.

2.3.2 Governance of climate data management systems

The CDMS governance component refers to a consistent set of WMO and national policies and governance processes needed to build a solid foundation for the establishment and management of authoritative and trustworthy sources of climate data and related services. Currently there are a number of WMO initiatives in train to establish consistent policies; this component summarises the best practices as they stand at the time of writing (2023), and may help to immediately improve many NMHS data management practices.

This component contains the following concepts:

- Data policy, including:
  - ✓ Organizational commitments;
  - ✓ Ensuring the viability and sustainability of CDMSs;
  - ✓ Intellectual property;
  - ✓ Data delivery;
  - ✓ Third-party data;
  - ✓ Climatological data management policy.

- Governance, including:
  - ✓ Adherence to broader WMO policies, including the WMO Data Policy;
  - ✓ Data governance;
  - ✓ ICT governance.

2.3.3 Data management

The Data management component addresses the functionality required to effectively manage climate data and includes the following concepts:

- Data ingest and extraction;
- Data rescue;
- Observations quality control;
- Dataset Quality assessment;
- Provenance and traceability of climate data;
- Management of climate metadata.

Metadata management is an important aspect of climate data management and is dealt in detail in chapters 4 and 5.

2.3.4 Data delivery

The Data delivery component refers to the functionality required to deliver climate data and includes the following concepts:

- Data discovery (both climate data and climate metadata);
- Data delivery in WMO formats;
- Data delivery based on open spatial standards (for example, the Open Geospatial Consortium (OGC) standards and the International Organization for Standardization (ISO) 19100 series).

2.3.5 Data analysis

The Data analysis component involves a wide variety of analytical techniques that are applied to climate data and may result in the generation of a range of derived data products. Some examples are:

- Techniques, including statistical, spatial and image analysis;
- Homogenization;
- Derived products, ranging from simple processes such as computation of relative humidity through to more complex products such as gridded analyses;
- Numerical modelling processes, specifically the ability to store the results of such processes within a CDMS and related infrastructure.

2.3.6 Data presentation

The Data presentation component represents a diverse set of techniques used to communicate climate-related information. These include:

- Written reports;
- Time-series climate data exploration via a graphical user interface with functionalities such as:
  - ✓ Generating a broad variety of business intelligence reports, including tables, graphs, scatter plots, histograms and ensembles;
  - ✓ Visualizing disparate data using, for example, cartographic techniques, diagrams and 3D;
  - ✓ Conducting an integrated search and dynamic exploration of disparate climate data and metadata using functionalities such as spatial intelligence techniques;
- Multimedia exploration of data via, for example, podcasts, videos or photographs;
- Allowing the download of viewed data via the graphical user interface.

2.3.7 ICT infrastructure

The ICT infrastructure components represent the functionalities required to support a CDMS.

2.4 Overview of functional components of a climate data management system

The schema below (Figure 2.4) has been included to enable readers to contextualize the major CDMS components outlined above and the specific CDMS components covered in detail throughout the remainder of this publication.

Readers may find this diagram useful for developing a comprehensive understanding of the detailed CDMS components.

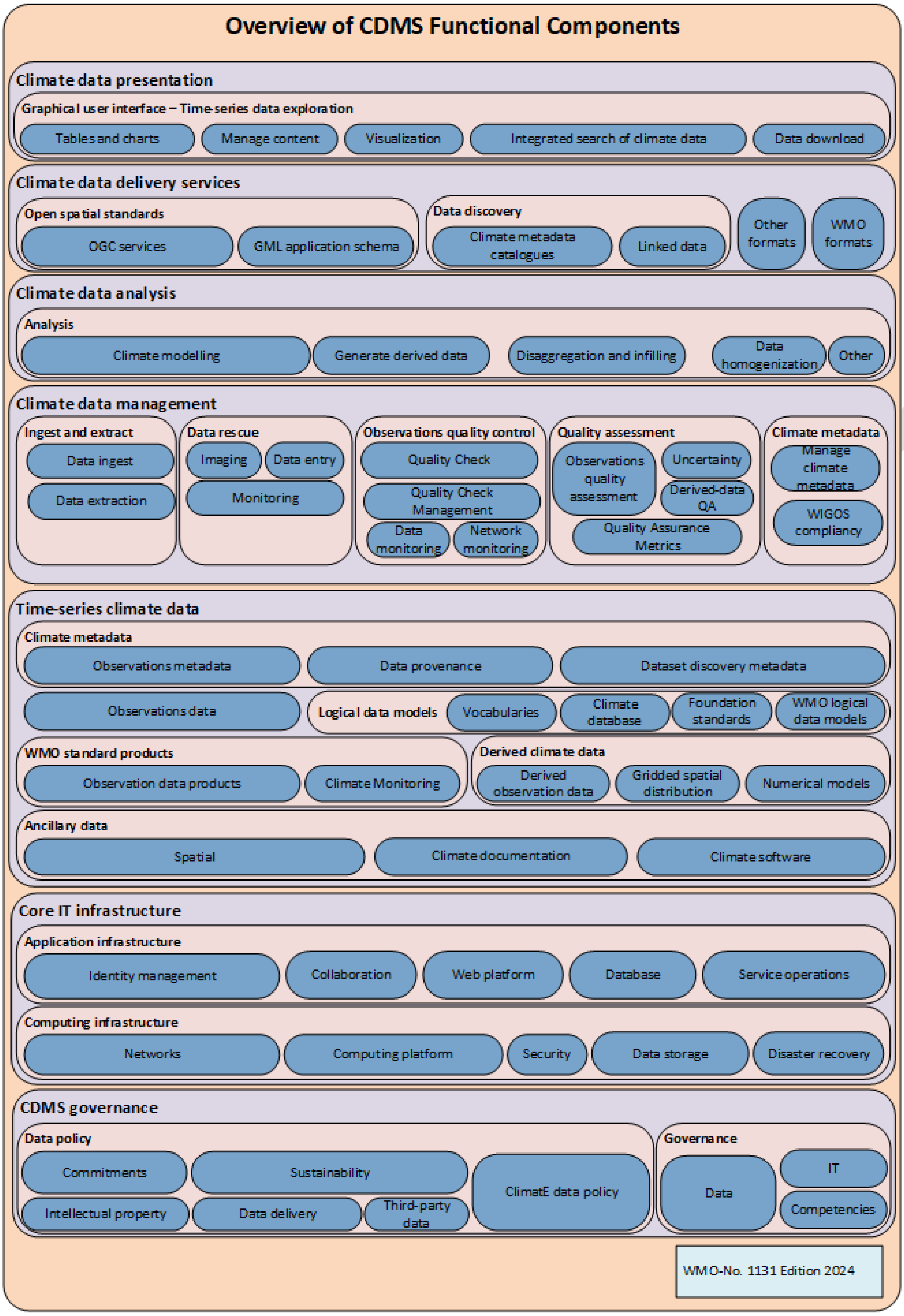


Figure 2.4. Overview of functional components of a climate data management system

2.5 Detailed exploration of the functional components of a climate data management system

Each of the major CDMS components discussed above is explored in more detail in the relevant sections of this publication, from Chapter 3 to Chapter 9 inclusive.

The components are deliberately discussed as abstract concepts, that is, we avoid recommending specific ICT solutions. A single component refers to a **functional requirement** that may be provided by a range of software and processes. Similarly, a single software application may provide the functionalities described in a number of components.

2.5.1 Note on applicability of the functional components of a climate data management system

2.5.1.1 Not all components are required

A key consideration that was used in preparing these specifications is that there was no difference in **basic** CDMS functionality for least developed, developing or developed countries.

A CDMS component classification scheme has been created to indicate whether a functionality is considered to be required, recommended or optional. This classification scheme and associated definitions are described in section 2.6.

NMHSs can use the CDMS component classification scheme to help them determine what functionality their organization should prioritise the development or acquisition of, and can afford to support over the long term.

In addition, NMHSs may find that CDMS functionalities (required, recommended, optional) may be available via other sources, such as Regional Climate Centres, the WMO Information System (WIS), or be accessible via Cloud Technologies or by partnering with other NMHSs.

2.5.1.2 Current and Future state of the CDMS

The detailed CDMS functional components described in this publication are wide-ranging and are intended to give readers an idea of the capability expected of CDMSs. This can be thought of as the future state of the CDMS.

It is recommended that NMHSs firstly assess how their current CDMS (if any), and related infrastructure, match the CDMS functional component requirements. This can be thought of as the current state of their CDMS.

NMHSs will then be able to determine the level of investment they can sustain in the long term to implement, support and maintain their CDMS functionality. This investment will not be restricted to just ICT hardware and software costs, but will also need to cover the specialist skills required to manage the CDMS. This includes expertise in climate data management, climate data analysis and a wide range of ICT skills.

See Chapter 10 for more considerations that may be relevant when implementing a CDMS.

Taking into account their current state, the amount of financial and human resources they can afford to invest, and the CDMS component classification scheme, NMHSs will then be able to plan their transition to the future state of their CDMS.

2.6 Component classification scheme

The following classification scheme is used in the detailed exploration of CDMS components to indicate whether a functionality is required, recommended or optional.

The classification can be used to assist NMHSs in determining and prioritizing what functionality they are able to implement.

Appendix 1 contains a complete set of the detailed component diagrams, which summarize the classification of the CDMS components.

Please note that the classifications given in subsequent Chapters have in some cases changed from the recommendations in the 2014 version of this publication, due to changes in priority of certain climate data requirements. They may change further with future revisions of this publication.

Required	Recommended	Optional

Note that this classification does not follow the wording convention used in the *WMO Technical Regulations* and regulatory annexes. If a “shall” in those WMO publications will certainly be a Required component in WMO-No. 1131, a “should” could become as well a Required component in WMO-No. 1131.

The Appendix 2 lists all components and their classifications. This list is also available in spreadsheet format (Excel) on the WMO e-Library and can be used to compare the classification of the components of an NMHS's climate data management system with the WMO criteria. This can be useful in identifying priorities for improvement of a CDMS, especially for data managers decision-makers.

The authors are well aware that some of the components noted as "Required" or "Recommended" may represent significant difficulties for some current systems that will have to comply with them. The "Required" or "Recommended" level indicates the level that a user of climate data should expect as state of the art at the time this document was written. The “Required” level is then more appropriate for the development of a new CDMS.

Required	The component's functionality is mandatory within a CDMS. There are many reasons why a component may be defined as mandatory, such as the need to comply with a specific WMO Technical Regulation or the *Guide to Climatological Practices* (WMO-No. 100).	*It is highly recommended that investment in the improvement of CDMS used by NMHSs and other climate data management entities prioritise these components.*
Recommended	This component's functionality is recommended in order to comply with current best practices.	Due to the evolving nature of the CDMS, these Recommended functions are becoming very useful and in some cases may be necessary, depending on the needs of the NMHS. It is possible that a component currently marked as Recommended may become Required in a future revision of the CDMS Specifications.
Optional	This component's functionality is optional. For example, an organization may wish to invest in more advanced climatological functionalities and processes in order to meet specific business needs.	Fulfilling these functions adds sophistication to the CDMS. A component may also be classified as optional if it is in the early stages of development within the WMO community.

As a final pieces of advice on required functionalities, the authors would like to highlight the two following points:

- Keep the climate record as safe as possible for as long as possible, and
- Take into high consideration the provenance of the data as explained in the chapter 4.3.3 of this publication.

2.7 Summary of changes from 2014 edition.

A review of developments in climate data management requirements in 2022 has resulted in the addition of a number of new components, and the reclassifying of some existing components (notably, Provenance metadata has been upgraded from “Optional” to “Required”). Such changes are reflected in the Component Summary diagram at the start of each section by the addition of indicators indicating whether the component is new (a little red square at the upper right corner of the box of the component) or if the component have changed from its former classification (green square if the component was “Recommended” and grey square if the component was “Optional” in the former edition.

2.8 References

2.8.1 WMO

#	Publisher	Title	N°	Edition
1.	WMO	*Technical Regulations*	WMO-No. 49	2019
2.	WMO	*Guide to Climatological Practices*	WMO-No. 100	2023
3.	WMO	*Manual on the WMO Information System*	WMO-No. 1060	2019
4.	WMO	*Manual on the WMO Integrated Global Observing System*	WMO-No. 1160	2021
5.	WMO	*Guide to the WMO Integrated Global Observing System*	WMO-No. 1165	2021
6.	WMO	*WIGOS Metadata Standard*	WMO No 1192	2019
7.	WMO	*Manual on the High Quality Global Data Management Framework for Climate*	WMO-No. 1238	2019

2.8.2 Links

Publisher	Title	URL	Current at

UCAR	Community Earth System Model	https://scied.ucar.edu/image/community-earth-system-model	Dec 2023

3 CDMS Governance

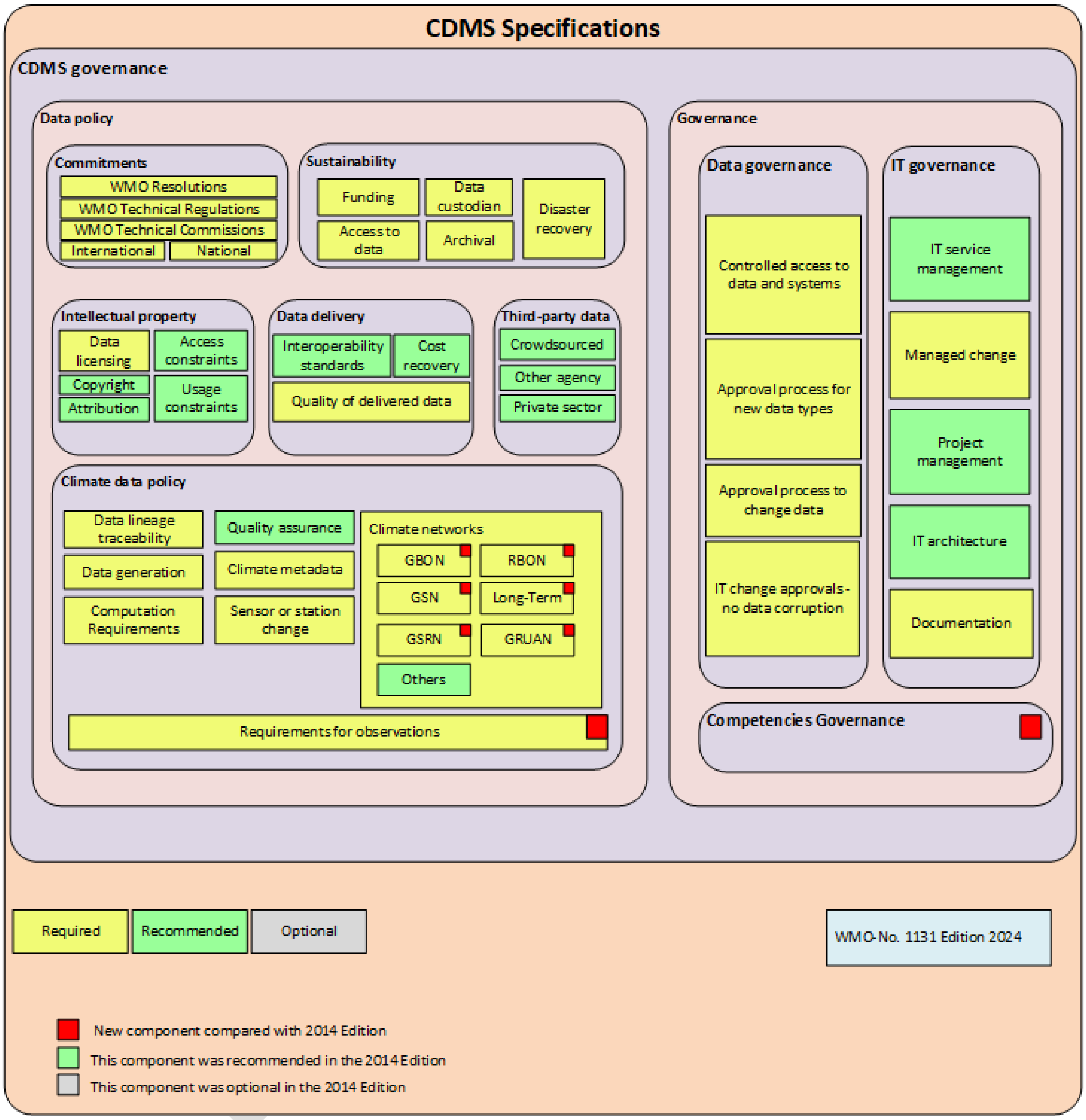


Figure 3.1. Overview on CDMS Governance components

One of the steps often overlooked when implementing a CDMS is to ensure that the NMHS has established a solid governance framework to protect the integrity of the climate record. This involves:

- Policies – the rules and general principles adopted by the organization to guide staff in climate data management activities.

- Governance – the processes, including approval processes, that have been defined and implemented to ensure adherence to CDMS-related policies.

Having such a foundation in place will save an organization considerable effort and will increase operational efficiency by ensuring that staff have a clear understanding of what is expected and permissible.

CDMS data management policy and governance is an area that requires considerable development to ensure consistent approaches to:

- Facilitate the establishment of authoritative data sources and related services;
- Enable data sharing and integration;
- Establish a clear understanding of intellectual property issues that apply to data;
- Establish a clear understanding of data access and usage constraints;
- Facilitate processes that ensure and enhance data integrity;
- Ensure the long-term sustainability of NMHS climate records.

In the interim, the components contained in this version of the CDMS Specifications will provide NMHSs with a starting point to begin preparing for issues that will arise.

3.1 Data policy

Note that a difference exists between WMO Data Policy, which is to do with the exchange of a wide range of data and products, including climate-related ones (Section 3.1.1.1), and data **management policies,** which are the subject of this Governance Chapter.

3.1.1 Commitments

What commitments has the implementing organization agreed to either explicitly through a contractual arrangement or implicitly through national policy agreements, regional and international agreements, WMO membership or other means?

3.1.1.1 WMO Resolutions	This component refers to the policy framework within WMO, governed at the highest level through Congress and the Executive Council. The list of current resolutions can be found in *Resolutions of Congress and the Executive Council* (WMO-No. 508). In 2021, Congress endorsed Resolution 1 (Cg-Ext 2021), the WMO Unified Policy for the International Exchange of Earth System Data which governs the exchange of data including climate data by Member countries, and replaces	Required

	a number of pre-existing WMO Resolutions on data exchange (Res No 25. 40 60 and 56 which are no longer in force). The major innovation in this resolution is the definition of core and recommended data including climate data. See WMO Unified Data Policy (Resolution 1). Annex 1 of the Resolution 1 lists the core and recommended data to be exchanged for the following Earth system domains: weather, climate, hydrology, atmospheric composition, cryosphere, oceans, and space weather. Annex 2 provides guidelines to Members on the application of the WMO Data Policy including recommended national practices regarding exchange of Earth system data. Annex 3 presents guidelines on the application of the data policy in public-private engagement. It describes also the need to respect national and regional legislation and policy, consistent with the principle of free and unrestricted data exchange. Annex 4 presents terms and definitions applicable to the rest of the Resolution. See: • *WMO Unified Data Policy*, 2022	
3.1.1.2 **WMO Technical Regulations**	This component covers standard practices and procedures, as well as recommended practices and procedures, for WMO Members to follow and implement. The main references are the three volumes of the WMO *Technical Regulations* (WMO-No. 49): • Volume I – General Meteorological Standards and Recommended Practices • Volume II – Meteorological Service for International Air Navigation • Volume III – Hydrology • Also see the publications of the annexes that are part of the Technical Regulations:	Required

	• Annex I – *International Cloud Atlas* (WMO-No. 407) • Annex II – *Manual on Codes* (WMO-No. 306) • Annex III – *Manual on the Global Telecommunication System* (WMO-No. 386) • Annex IV – *Manual on the Global Data-processing and Forecasting System* (WMO-No. 485) • Annex VI – *Manual on Marine Meteorological Services* (WMO-No. 558), • Annex VII – *Manual on the WMO Information System* (WMO-No. 1060) • Annex VIII – *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160) • Annex IX - *Manual on High Quality Global Data Management Framework for Climate* (WMO-No 1238)	
3.1.1.3 **WMO Technical Commissions**	This component includes the recommendations or best practices that are generally drawn up as guides published by the different WMO commissions: • Commission for Observation, Infrastructure and Information Systems (INFCOM) • Commission for Weather, Climate, Water and Related Environmental Services and Applications (SERCOM) • And the Research Board While most of the guides are relevant to CDMSs, it is recommended that particular attention be paid to: • *Guide to Instruments and Methods of Observation* (WMO-No. 8), • *Guide to Climatological Practices* (WMO-No. 100) • *Guide to Hydrological Practices Volume I: Hydrology form Measurement to Hydrological Information* (WMO-No 168)	Required

	• Guide to Marine Meteorological Services (WMO-No 471)	
3.1.1.4 **International**	This component concerns commitments that organizations are required to adhere to because they qualify as international agreements. Some examples are: • The Global Framework for Climate Services (GFCS), which has an important role in climate data management, climate monitoring and assessment, climate products and services, and climate information for adaptation and risk management • International data policy agreements, such as the Infrastructure for Spatial Information in Europe, which has defined and legislated common principles for European Union countries that enable the sharing of environmental spatial information among public-sector organizations. This generally includes climatological data • The Group on Earth Observations (GEO), which concerns the exchange of observations data • The OGC and ISO, which focus on the exchange of data via spatial standards. Note that not all standards are relevant	Required
3.1.1.5 **National**	This component represents national policies that organizations are required to comply with. Some examples are: • National government policies on open data and open-source software, records management (including retention policy) • National spatial data infrastructure frameworks that govern the exchange of national spatial data	Required

3.1.2 **Sustainability**

3.1.2.1 **Disaster recovery**	This component refers to a policy that governs how an organization implements disaster recovery and business continuity solutions and covers issues such as how to handle data backups. Some initial issues to consider are:
	• How important is the climate record?
	  ✓ How often should climate data be backed up? While the [Wikipedia article on backups]() can provide some background information, this question must be answered by each NMHS. Their answer will need to take the following into account:
	  ✓ How important are their data?
	  ✓ How much data can they afford to lose?
	  ✓ How much work can they afford to redo?
	  ✓ Can the data be regenerated?
	  ✓ In the interim, while NMHSs are assessing their requirements, it is suggested that daily backups be made with weekly off-site storage.
	• How many redundant copies of the data should be kept?
	• What secure off-site facility should be used to store disaster recovery versions of the data?
	• What is the acceptable downtime for each of the key CDMS applications?
	• How high is the risk of catastrophic damage to the organization's infrastructure due to a range of factors, including:
	  ✓ Natural disasters, such as flooding, bush fires, tidal waves, cyclones or earthquakes?
	  ✓ Man-made disturbances, such as military conflict, terrorism and fires?
	• Where should the secure off-site storage be located? For NMHSs situated in areas where their infrastructure is likely to be damaged, it may be appropriate to place
		Required

	the secure off-site storage in another city or country • Is it appropriate to use cloud-based services to store and manage climate data and offer CDMS services to end-users? It is to be emphasized that irreplaceable original observations and their metadata must be retained in perpetuity.	
3.1.2.2 **Funding**	This component concerns a policy that regulates how an organization funds its CDMS to ensure that it is sustainable. This includes sufficient funding for: • Climate data management specialists, including staff with knowledge of observations data, data rescue, quality assurance, etc. • IT specialists who support the everyday maintenance of CDMS applications and related databases and IT equipment • IT specialists who conduct enhancements to the CDMS and related IT environments • The provision and scheduled upgrade of IT systems to ensure that the appropriate IT environment has been implemented to support the CDMS • Before new data types are added funding must be available both to create the necessary storage environment and maintain it • Regular staff training	Required
3.1.2.3 **Data custodian**	This component represents a policy that governs how an organization manages and maintains its climate data A data custodian is typically a senior-level manager who is accountable for the integrity of the climate record of the NMHS Some examples of a data custodian's duties are: • Preserving the integrity of the climate record, including quality control and ensuring that observational networks	Required

	provide data that are suitable for climate purposes • Championing the cause of data management to ensure that sufficient funding is allocated and managed effectively so that a reliable climate record remains viable • Facilitating the development and maintenance of suitable policies governing climate data • Ensuring that climate data are effectively managed and maintained • Ensuring that observations metadata, discovery metadata and data provenance are effectively maintained • Formally delegating authority to appropriate staff members, together with related performance accountabilities • Taking primary responsibility for CDMS applications, CDMS enhancement and IT maintenance projects • Ensuring that IT changes do not corrupt the climate record • Implementing and monitoring relevant key performance indicators to help monitor ongoing performance of the CDMS and related processes	
3.1.2.4 **Access to data**	This component involves a policy that governs how an organization allows access to data Some issues to consider are: • How is read-and-write access to climate data controlled? • What training is needed before staff are granted write access? • What types of controls are needed to monitor and approve changes to the climate record to facilitate data maintenance and error rectification? • How are any specific access constraints, such as security, contractual or commercial constraints, to be managed?	Required

	Note: The issue of open data is increasing in relevance for many national governments that have made open-data commitments. Therefore, organizations should have a clear understanding of, and a clear policy for, the provision of NMHS climate data in accordance with open data principles (see Wikipedia article on open data). In summary, open data is about ensuring that data are freely available for reuse with no constraints apart from the requirement to acknowledge the data source and/or ensure that the data are made available under a share-alike or similar agreement.	
3.1.2.5 Archival policy	This component represents a policy that dictates how organizations archive their climate data, including both digital and hard-copy historical records. This archive should be considered as permanent to ensure that the climate record is available for use by future generations. See the definition of “climate record” near the beginning of the document which incorporates relevant metadata. Therefore, care should be taken to ensure that the data are preserved in a format that users will be able to access and use in years to come.	Required

1

3.1.3 Intellectual property

3.1.3.1 Data licensing	This component refers to policies that ensure that any data licensing agreements that relate to the use of a dataset are clearly accessible and understood. For example: • What data licences apply to NMHS data? • What data licences will the NMHS release its data under? For example:	Required

<table>
<tr><td></td><td>Is the NMHS part of a governmental Open Data policy?
✓ Is a Creative Commons licence appropriate? (See Creative Commons website)
What data licences used by external data-providers are permitted for use within the NMHS?
Can the data be distributed to third parties?
Will the NMHS use and/or archive data that are not covered by an appropriate data licence?
Will the NMHS comply with the licensing of data passed to it, or will it choose not to use or archive the data?</td><td></td></tr>
<tr><td>3.1.3.2 Access constraints</td><td>This component concerns policies that clearly define any access constraints relating to the use of climate data. This would be particularly pertinent to data held on behalf of third parties.

For example:
What may the NMHS do with the data?
Can the data be used for the organization’s website, web services, publications and so forth?
Do any commercial or contractual constraints apply to the use of the data?
Are the data subject to any national security constraints?</td><td>Recommended</td></tr>
<tr><td>3.1.3.3 Usage constraints</td><td>This component covers policies that ensure that any constraints imposed on the end-user regarding the use of a dataset are clearly understood. These constraints may apply to the NMHS or to third parties.

For example:
Can the data be freely reused?
Is there a cost that applies to the use of the data? See the Commercial component (3.1.5.3) for a number of related considerations.</td><td>Recommended</td></tr>
</table>

	• Can derived products be made using the data? • Can the data be used for commercial purposes? • Can the data be used for private, study or research purposes? • Can the data be shared with others? The Creative Commons website provides examples of data usage constraints and related licences.	
3.1.3.4 **Copyright**	This component refers to policies that clearly explain any copyright issues relating to the use of climate data.	Recommended
3.1.3.5 **Attribution**	This component deals with policies that ensure that any data attribution issues relating to the use of climate data are clearly understood. For example: • Should the source of the data be acknowledged? • How does the NMHS ensure that any attribution text required by the data provider is applied when the data are used?	Recommended

1

3.1.4 **Data delivery**

3.1.4.1 **Interoperability standards**	This component involves policies that ensure that climate data are delivered using appropriate open interoperability standards, such as open spatial standards. This will ensure that data are represented in a form that facilitates data interoperability and are accessible to a wide range of end-users from disparate industries using a wide variety of proprietary and open-source software applications. Note: It is possible to enforce the use of data formats such as BUFR or GRIB within an NMHS in an attempt to facilitate interoperability. However,	Recommended

	in practice, only NMHSs ad closely affiliated organizations may possess the capacity to comprehend the formats and the necessary software to utilize the data. For more information, see section 8.1 of this publication.	
3.1.4.2 **Quality of delivered data**	This component refers to policies that clearly define the issues relating to the quality of the climate data delivered by NMHSs. For example: • What level of quality of climate data is the organization committed to deliver and under what conditions? • Will the organization only provide quality-controlled and/or homogenized data? • Will the organization provide raw observations? • When is it appropriate to deliver data at each quality level?	Required
3.1.4.3 **Cost recovery**	This component refers to policies that ensure that issues relating to the recovery of costs for the provision of NMHS climate data services are clearly understood and communicated. These policies should also take into account: • NMHS commitments to WMO relating to Resolutions 1, the WMO Unified Policy for the International Exchange of Earth System Data, as discussed above in 3.1.1.1. • National commitments to open data policies.	Recommended

1

3.1.5 Third-party data		
3.1.5.1 **Crowdsourced**	This component refers to policies that address and explain issues relating to the use of crowdsourced climate data by the NMHS (see [Wikipedia article on crowdsourcing]).	Recommended

	In summary, crowdsourcing is about obtaining data from a large pool of volunteers who would like to collaborate with the NMHS or other organizations.
	
	Crowdsourcing has considerable potential for enhancing data generated by NMHSs, provided that the data are used appropriately. There are many examples of crowdsourcing initiatives that have generated very useful data. Some examples are:
	• OpenStreetMap
	• Old Weather – a data rescue project to digitize meteorological observations from old ship logs
	• Weather Observations Website (WOW), Met Office (United Kingdom)
	• Veilleurs du temps, Meteo-France
	
	Some issues that crowdsourcing policies should consider are:
	• Only accepting crowdsourced data when the supplier has agreed to either provide the data under a Creative Commons ShareAlike licence (see Creative Commons ShareAlike page) or to have its intellectual property rights assigned to the NMHS
	• Having intellectual property rights to the data recorded for future use
	• When the data contributed are observations data:
	  ✓ Having the contributor commit to providing and maintaining suitable observations metadata for their station(s) and sensor(s)
	  ✓ Having the contributor agree to appropriate site visits by NMHS staff to assist with quality assurance processes
	• Deciding when it is appropriate to use crowdsourced data and under what conditions
	• Deciding how the crowdsourced data will be managed	

	NMHSs should develop a proper data management plan that addresses all of the issues raised in this component.	
3.1.5.2 Other agency	This component refers to policies that provide clear explanations of issues regarding the use of climate-related data captured and maintained by government agencies external to the NMHS. The agreements between the parties should be documented. Examples of issues to consider include: • Is there a clear data licence allowing the NMHS to use the data for any relevant purpose, including for internal use, to create derived products, to publish the data on the NMHS website or for data redistribution, if required? • Are there any costs associated with the use of the data? • Is it a high-quality dataset? Do the discovery metadata clearly describe the intended use, lineage and quality assessment of the data? • What constraints apply to the use of the data? • If the data are observations data, are there high-quality observation metadata available to support CDMS activities? • If the data are of similar quality to NMHS data, share similar functions and help enhance the NMHS data, is the provider willing to consider joint ownership and maintenance of the data?	Recommended
3.1.5.3 Private sector	This component refers to policies that address and clarify issues on the use of climate-related data captured and maintained by the private sectors. In addition to the considerations discussed under the Other agency component (3.1.5.2), other issues that commercial policies should consider include: • What contractual arrangements apply to the use of the data?	Recommended

	• What costs apply explicitly to the use of the data? • Are there constraints on the number of users who can access the data at the same time? • Are there pricing qualifications based on the specific server environment that hosts the data, or more specifically on: ✓ The number of central processing units, processor cores, threads and so forth? ✓ Whether there are any restrictions based on the virtualization of the server? ✓ Is there a pricing model based on the number of views of the data within an end user's web browser, or similar application? • What are time constraints? (E.g. embargo period and data availability duration.) What must be done with the data at the end of data availability period if applicable?? If so, what must be done with the data and the products generated using these data at the end of this period? • Are the data only available via a subscription service or web service? If so, what impact will this have on NMHS operations in the event of a disruption of the service or when an invoice is not paid on time? • Are there any explicit commercial constraints on the use of the data? • Are the data actively maintained? • Do the costs, constraints and associated risks call for an arrangement to be made between the NMHS and the data provider? • Are there alternative datasets of similar functionality and quality that could be obtained elsewhere, such as a community-maintained dataset like OpenStreetMap?	

1

3.1.6 Climate data policy

3.1.6.1 **Climate metadata**	This component refers to policies that ensure that appropriate climate metadata are maintained to facilitate a better understanding of climate data. As defined in section 4.3, climate metadata include metadata on observations, discovery and data provenance.	Required
3.1.6.2 **Observations Metadata**	This component concerns policies about observations metadata (mandatory, conditional and optional) compatible with the WIGOS Metadata Standard that WMO members shall make available internationally. The metadata on observations are made available through the database of the Observing Systems Capability Analysis and Review (OSCAR) tool. See: • The Manual on the *WMO Integrated Global Observing System* (WMO-No. 1160)	Required
3.1.6.3 **Data lineage traceability**	This component concerns policies ensuring that the CDMS is able to trace the lineage of climate data from published scientific texts and other papers, homogenised datasets, etc. back to raw observations. May be seen as a pillar to ensure reproducibility, helping to build trust and credibility in the climate rerord. This will include the ability to reproduce specific data that were held in the climate database at a particular point in time. Note that it may not be practical to implement this policy retrospectively, i.e. for papers published in the past. This specific policy requirement has become increasingly relevant following the so-called Climategate issue. One of the conclusions of the UK parliamentary enquiry that investigated this issue was that: It is not standard practice in climate science to publish the raw data and the computer code in academic papers. However, climate science is a matter of great importance and the quality of the science should be	Required

	irreproachable. We therefore consider that climate scientists should take steps to make available all the data that support their work (including raw data) and full methodological workings (including the computer codes). Had both been available, many of the problems at UEA [University of East Anglia] could have been avoided. (UK Parliament Science and Technology Committee, 2011)	
3.1.6.4 Data generation	This component covers a range of policies that govern the generation and interpretation of observation variables. As these rules have changed and may change again in the future, the policies will need to cover past, current and future data generation policies. Some considerations are: • These policies will need to cover the methods, algorithms, models and software source code used to generate data. • They should include tuning for national climatological day, climatological hour etc. The definition of climatological days may differ: ✓ Between NMHS ✓ Between different stations within a single NMHS ✓ Over time within the same NMHS • Between a given NMHS policy and the method applied by the software used in automatic weather stations (which often appear as a black box to NMHSs). For example, a climatological day could represent the following time periods: ✓ From 0000 LST to 2300 LST ✓ From 0100 LST to 0000 LST the next day ✓ From 2346 LST the previous day to 2345 LST the current day ✓ From 0900 LST the previous day to 0900 LST the current day LST: Local Standard Time Those policies should also cover the rules relating to the management of missing observations.	Required

	The CDMS should be flexible enough to enable NMHSs to readily incorporate their national practices in order to avoid introducing bias in data time-series. Some other specific arithmetical characteristics influence the data computation and should be decided for the system. Such as the data precision throughout a whole process, the rounding method used, etc. For example, the algorithm for any data computation shall be well described with its data precision, its rounding method and formula. See the *WMO Guidelines on the calculation of climate normals* (WMO-No. 1203) at the paragraph 4.7 on Data precision and rounding where some rounding methods are mentioned ("Ties to odd" and "Ties to even"). Also see the WMO formulae for the meteorological variables such as pressure, humidity in the *Guide to Instruments and Methods of Observation*, Volume I - Measurement of Meteorological Variables (WMO-No. 8)	
3.1.6.5 **Climate networks**	This component concerns a range of policies that determine the design of observation networks including for climate application and establish the station and network operations, with observation times, on-site quality control, observer training, station inspections, etc. For more information, see: • *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160) • *Guide to the WMO Integrated Global Observing System* (WMO-No. 1165) • *Guide to Climatological Practices* (WMO-No. 100) • *Guide to agricultural meteorological practices* (WMO-No. 134, Updated 2012) • *Guide to Hydrological Practices* (WMO-No. 168, 2009) The CDMS should be in capacity to control the assignation of stations to particular networks within	Required

their inherent constraints. The following networks are required:

- The GBON, the Global Basic Observing Network of WMO, contributes to meeting the requirements of Global Numerical Weather Prediction, including reanalysis in support of climate monitoring. For example, the point 3.2.2.7 in WMO-No. 1160 says that: “Members shall maintain the continuous operation of a set of surface land stations/platforms that observe, at a minimum, atmospheric pressure, air temperature, humidity, horizontal wind, precipitation and snow depth, where applicable, located such that GBON has a horizontal resolution of 200 km or higher for all of these variables, with an hourly frequency”
- The RBON, the Regional Basic Synoptic Network of WMO, should comply with constraints on sustainability (e.g. committed to at least a ten-year period of operation), with a horizontal resolution generally round 100 km and with international exchange in real time or near-real time
- The GSN, the Global Climate Observing System (GCOS) Surface Network, and the GUAN, the Global Climate Observing System (GCOS) Upper-Air Network should both follow specifications described in WMO-N. 1160
- The Global Climate Observing System (GCOS) Reference Upper-air Network (GRUAN) together with the Global Climate Observing System (GCOS) Surface Reference Network (GSRN). GRUAN and GSRN should provide long-term high-quality climate records
- The WMO long-term observing stations network, centennial and 75+years, have also some requirements that should be known by data managers and so be defined into the system. See *Centennial Observing Stations* (WMO-No. 1296).

At international, regional and national level, some other climate networks may have specific constraints different from those of WMO and from those of the GCOS.

3.1.6.6 Requirements for observations	This component reflects the requirements on observations that should be applied per physical variable and per domain such as aeronautic, agriculture, hydrology, climate monitoring, etc.. Those requirements are managed and available within the WMO OSCAR/Requirements database. Those requirements answer to a performance level quantified in terms of up to six criteria: uncertainty, horizontal resolution, vertical resolution, observing cycle, timeliness and stability. For each of these criteria the database indicates 3 values: • The **"threshold"** is the minimum requirement to be met to ensure that data are useful • The **"goal"** is an ideal requirement above which further improvements are not necessary • The **"breakthrough"** is an intermediate level between "threshold" and "goal" which, if achieved, would result in a significant improvement for the targeted application. The breakthrough level may be considered as an optimum, from a cost-benefit point of view, when planning or designing observing systems. See the OSCAR/Requirements database at https://space.oscar.wmo.int/requirements	Required
3.1.6.7 Sensor or station change	This component covers a range of policies that apply to changes affecting a station or sensor (such as a replacement or relocation). These policies are of great importance for time-series analysis. They have implications for climate metadata and for the possible implementation of parallel observations for certain time periods._These are vital to ensure effective homogenization of data time-series. For more information, see: • *Guide to the Global Observing System* (WMO-No. 488) *If previous long-standing climatological data series have to be extended in time by data provided by automatic synoptic stations, parallel measurements by conventional and automatic observing methods are indispensable to ensure continuity of the records. One year of parallel*	Required

	measurements does not suffice; preference is given to at least two years, depending on the climatic region. • *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), Volume III- Observing Systems: The following general guidelines are suggested for a sufficient operational overlap between existing and new automated systems: (i) Wind speed and direction: 12 months (ii) Temperature, humidity, sunshine, evaporation: 24 months (iii) Precipitation: 60 months. (It will often be advantageous to have an ombrometer operated in parallel with the automatic raingauge.) A useful compromise would be an overlap period of 24 months (i.e. two seasonal cycles). • *Guideline on Homogenization* (WMO-No. 1245, 2020) • *Challenges in the Transition from Conventional to Automatic Meteorological Observing Networks for Long-term Climate Record*, (WMO-No 1202, 2017)	
3.1.6.8 **Quality assurance**	This component refers to policies that ensure that quality assurance issues for climate data within an organization are clearly understood. Some issues to consider are: • What level of quality assurance (including quality control) can the organization afford to maintain over the long term? • Will the organization conduct quality assurance/control checks on all observation phenomena or only on a subset? • What quality assurance levels (or tiers) are used by the organization for the long term? • How is each quality assurance tier defined? • What quality assurance tests must data successfully pass before they are promoted to the next tier? • How could the ISO 9000 series of quality management standards help improve data management processes?	Recommended

	See: • *Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers* (2017 edition; WMO-No. 1100) • *Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications* (WMO-No.1269)	

1

3.2 Governance

3.2.1 Data governance

3.2.1.1 Controlled access to data and systems	This component refers to governance processes that ensure a clear understanding of user access to data and IT systems within an organization. Some data-related issues to consider are: • Which staff roles have read or read-and-write access to each type of data? If appropriate, to which subset of stations, of variables, etc? • Which roles have read or read-and-write access to each quality tier of data? • What process is in place for designating staff to each role? • What process applies to ensure that access to data is subject to approval by a delegate of the data custodian? • Should staff who have write access to data stored within a database only be able to change data under software control? • Should each successful change to observations data be audited to ensure that the change, the details of the operator who made the change	Required

	and the time of the change have all been recorded? Some system-related issues to consider are: • What roles should be given to run-time users to access data? • Which roles have the ability to change operational software or systems? • What process applies to ensure that access to systems and software applications is subject to approval by a delegate of the data custodian?	
3.2.1.2 **Approval process for new data types**	This component refers to governance processes that ensure that issues relating to the approval of new data types within an organization are clearly understood. Some issues to consider are: • What needs to be considered prior to accepting a new data type for long-term archival (such as user requirements, scientific requirements or a new statutory requirement)? • What are the projected storage requirements for the data? • What is the appropriate format for storing the data over the long term? • Is there likely to be an application that can read that format in 10 or 20 years? • What skills are available to process and analyse the data over the long term? • Is suitable funding available to maintain the data over the long term? • Will cloud storage be used to host the data, and are there policies in place dictating how the data may be used? • Strongly recommended that suggestions to incorporate new data types are regulated by the need to develop a Data (or Information) Management Plan, as defined in the	Required

	WMO "Guidance on Information Management" document	
3.2.1.3 Approval process to change data	This component refers to governance processes that clearly define the approval process required to modify the data held within an organization's climate record.	Required
3.2.1.4 IT change approvals – no data corruption	This component refers to governance processes that ensure that any IT change does not result in an unexpected change to, loss of or corruption of the climate record.	Required

1

3.2.2 IT governance

This collection of components refers to the overall governance of information technology to ensure that effective CDMSs are developed, enhanced and maintained.

While these components are specifically aimed at larger organizations with a substantial investment in information and communication technology, smaller organizations can also benefit from investment in the types of issues discussed.

These components are also relevant to broader initiatives, such as when an aid or development agency sponsors CDMS development and implementation.

These issues will only be discussed very briefly, as each component refers to a substantial body of knowledge that will require ongoing investment from specialists to ensure effective management of resources and funds for information and communication technology.

3.2.2.1 IT service management	This component refers to an overarching framework for IT service management used to ensure effective, efficient and cost-effective management of the delivery of business value from IT services. This framework is typically based on global best practices. It comprises a consistent framework within which the remaining IT governance components can operate in a tightly integrated manner.	Recommended

	For more information, see: • Wikipedia overview of IT service management • Information Technology Infrastructure Library (ITIL) •	
3.2.2.2 **Managed change**	This component covers governance processes that ensure that any change to the CDMS is carefully managed. Uncontrolled change can result in chaos for users of the CDMS, for example: • Systems may break down for no known reason. • Data corruptions or even data loss may occur and may not be detected at all. • Provenance of data is severely impacted, as data managers do not know what algorithms or processes have been applied to the data. This shows the importance of keeping provenance data to a large extent not locked into the IT vendor's platform solution and its internal representations. Long term availability and understanding of provenance data should be facilitated by the adoption of provenance models and management software that are interoperable and platform independent, respectively. • Considerable disruption may be experienced by staff that rely on the CDMS for day-to-day activities and may inconvenience users who rely on the data and derived products. • Data corruption, data loss and lack of availability of key systems may have a significant impact on the reputation of the NMHS. Some issues to consider: • How is the change process managed?	Required

	• Who needs to be consulted when considering a potential change? • How is the proposed change analysed to assess its potential impact on data integrity and on other CDMS components? • Who can authorize a change? • What testing is required to mitigate the potential impacts of the change and give confidence that the change is desirable? • How can the change be made with minimal impact? • Are there any dependencies between a series of proposed changes that may dictate that the changes occur in a particular order? • How is the change to be made and tested? • If the change fails, how can the system be rolled back to its previous state? • How can CDMS components be implemented such that they are self-testing to support concepts such as continuous integration? • Will processes based on an IT service management framework such as ITIL help improve the management of IT-related changes? (See Wikipedia article on IT service management)	
3.2.2.3 **Project management**	This component concerns governance processes that ensure that any development activity, infrastructure change or other enhancement related to the CDMS is carefully managed. Good project managers and project governance processes can mean the difference between a project delivering desired results or the same project failing at great expense, frustrating users and possibly corrupting the climate record. Uncontrolled enhancement can result in:	Recommended

	• An undesirable functionality added to the CDMS • Lower-priority work undertaken at the expense of higher-priority tasks • The development of activities that do not have a clearly defined scope, set of deliverables, timeline and budget Some additional considerations include: • How is the business benefit for a proposed project defined and assessed? • How are the relative priorities of projects assessed? • How are projects managed? What approach is adopted? • What process is used to monitor the progress of a project? • How are the project deliverables assessed to ensure that they meet requirements? • What lessons can be learned from a finished project?	
3.2.2.4 IT architecture	This component refers to strategic IT governance processes that ensure that CDMSs and related IT systems are carefully designed so that the science and NMHS requirements for CDMSs may be effectively and efficiently implemented. Uncontrolled development of components can have adverse impacts on the CDMS, including on the ability of the NMHS to maintain the CDMS over the long term. For example: • Reliance on proprietary solutions provided by a single vendor could result in a situation known as vendor lock-in. As time passes, more and more components are developed based on the single solution. It can be very difficult and costly for a NMHS to move to a new CDMS-related solution if the current vendor’s product becomes unaffordable or if the vendor	Recommended

	goes out of business, decides that it is no longer in its interest to offer CDMS components, ceases to maintain its systems or no longer offers a competitive CDMS solution • Similarly, uncontrolled bespoke development of CDMS components by NMHS staff or contractors can also be a problem. This may result in a wide variety of disparate CDMS components that cannot be easily integrated. Many different types of technology could be used, resulting in high maintenance costs. Key-person dependencies could also develop, leaving the NMHS with a significant and costly issue to resolve, particularly if the key person leaves the organization and the CDMS ceases to operate • There is a risk that CDMS component solutions could be developed that do not use climate data effectively, for example: ✓ Data may be replicated from the climate database and made into stand-alone files to support a particular technology or due to the developer’s lack of experience. Whereas the data in the climate database are subject to a lengthy quality assurance lifecycle and may change, the replicated data are fixed at a point in time and may not be updated to reflect later changes to the climate record. This may become a significant issue for the NMHS, particularly if this practice is systematically applied by many developers, across many applications and over many years ✓ The data extraction process may not take into account quality assurance flags and just present raw observations with no	

	indication of data quality or reliability ✓ Developers may not understand the complexities of the climate database data model and extract incorrect data ✓ Components may have any number of data inconsistency or misuse issues. What is the impact on the reputation of the NMHS if the same data are presented with differing values by different CDMS components? There may also be architectural considerations that are better off understood early in the CDMS development so that governance processes can ensure that the considerations are implemented within developed and/or implemented software One issue to consider is whether the CDMS should be able to work in multiple languages both in term of the user interface as well as any text content of the various information managed by the CDMS. This would be particularly important in countries that speak multiple languages. The answer to this will have implications for the design of software for the user interface of CDMS components, as well as for the generation and presentation of data products IT architecture is a specialized field of information technology that is typically undertaken by highly experienced professionals with very broad IT experience. Ideally, this task will be conducted by experienced NMHS staff to ensure that consistent CDMS components are implemented in accordance with a strategic vision. See also Wikipedia articles on: • Enterprise architecture • Solution architecture	

1

3.2.2.5 **Documentation**	This component covers governance processes that ensure that the CDMS is adequately documented and that this documentation is kept up to date to facilitate efficient day-to-day use by staff and ease the learning curve of new staff, contractors, consultants, etc. This documentation is very broad and includes: • An overview of the CDMS • An overview of the data being managed both within and outside the climate database • The CDMS components, design, business requirements, architecture, test plans and deployment processes • CDMS policies and governance processes • CDMS backup and disaster recovery processes • IT systems management and administration processes • Various CDMS-related metrics	Required

3.2.3 Competencies governance

This component establishes training and education programs for the personnel of NMHSs in order for them to acquire the necessary competencies to carry on the different tasks presented in this publication.

For further information on the establishment of competency and training programs for climate services in general, including in climate data management, see:

- *Guidelines on the Role, Operations and Management of National Meteorological and Hydrological Services* (WMO-No 1195, 2017)
- *Guidelines on Quality Management in Climate Services (WMO-No 1221, 2018)*
- *Guide to Competency* (WMO-No 1205, 2018)
- *Capacity Development for Climate Services: Guidelines for National Meteorological and Hydrological Services* (WMO-No 1247, 2020)
- *Guidelines for the Assessment of Competencies for Provision of Climate Services* (WMO-No. 1285, 2022)

- *Basic instructional package for climate services* (in publication process during the edition of this document)
- *Guide to the Implementation of Education and Training Standards in Meteorology and Hydrology*, volume I – Meteorology, (WMO-No.1083, 2015)
- *Compendium of WMO Competency Frameworks* (WMO-No. 1209, 2019)

3.2.3.1 Competencies governance		Optional

3.3 References

3.3.1 WMO

Title	WMO #	Publication
WMO Unified Data Policy	NA	2022
Basic instructional package for climate services	NA	
A Proposed Long Term Resource Plan for Climate Data Management and Rescue Activities	NA	2017
Guide to Instruments and Methods of Observation	WMO-No. 8	2021
Technical Regulations Basic Documents No. 2	WMO-No. 49	2019
Guide to Climatological Practices	WMO-No. 100	2023
Guide to Agricultural Meteorological Practices	WMO-No. 134	2010
Guide to Hydrological Practices Vol 1: From Measurement to Hydrological information	WMO-No. 168	2008
Manual on the Global Telecommunication System	WMO-No. 386	2015
International Cloud Atlas: Manual on the Observation of Clouds and Other Meteors	WMO-No. 407	2017
Guide to Marine Meteorological Services	WMO-No. 471	2018
Manual on the Global Data-processing and Forecasting System	WMO-No. 485	2019
Guide to the Global Observing System	WMO-No. 488	2013
Resolutions and Decisions of Congress and the Executive Council	WMO-No. 508	2020
Manual on Marine Meteorological Services	WMO-No. 558	2018

Guide to the Implementation of Education and Training Standards in Meteorology and Hydrology	WMO-No. 1083	2015
Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers	WMO-No. 1100	2017
Manual on the WMO Integrated Global Observing System	WMO-No. 1160	2021
Guide to the WMO Integrated Global Observing System	WMO-No. 1165	2021
Guidelines on the Role, Operation and Management of National Meteorological and Hydrological Services	WMO-No. 1195	2017
Challenges in the transition to Automatic Meteorological Observing Networks for long-term climate records	WMO-No. 1202	2017
WMO Guidelines on the calculation of climate normals	WMO-No. 1203	2019
Guide to Competency	WMO-No. 1205	2018
Compendium of WMO Competency Frameworks	WMO-No. 1209	2019
Guidelines on Quality Management in Climate Services	WMO-No. 1221	2018
Manual on the High-Quality Data Management Framework for Climate	WMO-No. 1238	2019
Guidelines on homogenisation	WMO-No. 1245	2020
Capacity Development for Climate Services: Guidelines for National Meteorological and Hydrological Services	WMO-No. 1247	2020
WMO Guidelines on surface station climate data Quality Control and Quality Assurance for climate applications	WMO-No. 1269	2021
Guidelines for the Assessment of Competencies for Provision of Climate Services	WMO-No. 1285	2022
Centennial Observing Stations	WMO-No. 1296	2022

3.3.2 Links

Publisher	Title	URL	Current at
Creative Commons	Licences	https://creativecommons.org/licenses/	Dec. 2023
Creative Commons	Attribution ShareAlike	https://creativecommons.org/licenses/by-sa/3.0/deed.en_US	Dec. 2023
ITIL	Information Technology Infrastructure Library	https://www.axelos.com/certifications/itil-service-management/	Dec. 2023
Wikipedia	Backup	https://en.wikipedia.org/wiki/Backup	Dec. 2023
Wikipedia	Open Data	http://en.wikipedia.org/wiki/Open_data	Dec. 2023
Wikipedia	crowdsourcing	http://en.wikipedia.org/wiki/Crowdsourcing	Dec. 2023
Wikipedia	ShareAlike	Creative Commons http://creativecommons.org/licenses/by-sa/3.0/deed.en_USpage	Dec. 2023
Wikipedia	Enterprise architecture	https://en.wikipedia.org/wiki/Enterprise_architecture	Dec. 2023
Wikipedia	Solution architecture	http://en.wikipedia.org/wiki/Solution_architecture	Dec. 2023
WMO	OSCAR Requirements	https://space.oscar.wmo.int/requirements	Dec. 2023

1

4 Time-series climate data

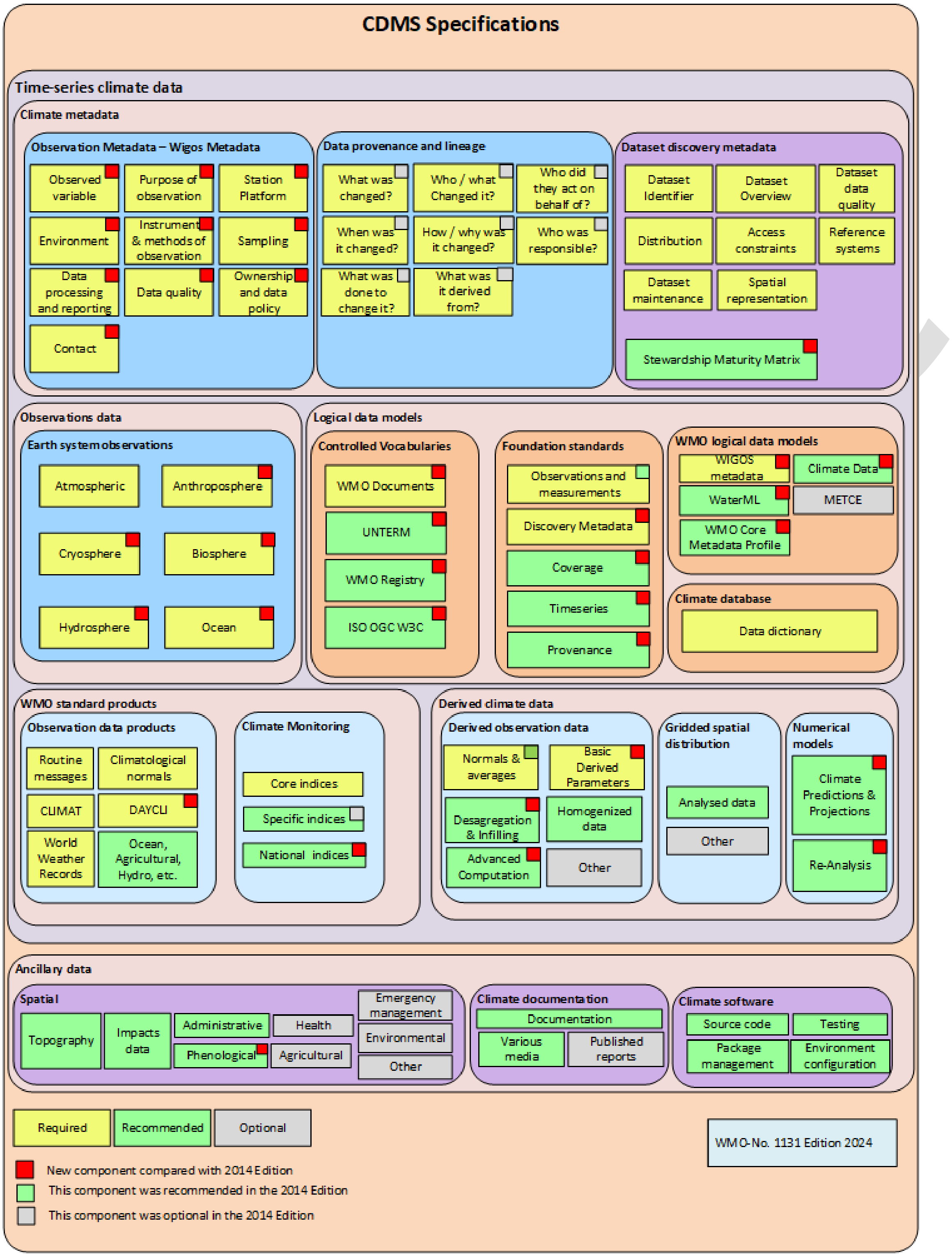


Figure 4.1 Time-series climate data components

The definition of what is considered to be Climate Data has been broadened beyond traditional meteorological observations with this CDMS Specifications to ensure that consideration is given to a wide range of data and information that is relevant to climate and related uses.

An attempt at a definition could be as follows: “Past, present and future data – observational, instrumental, proxy, derived, or value-added – which are utilised to enable climate services and research.”

Therefore, the definition of what is data of interest for climate purposes has been extended to include:

- Atmospheric, anthropospheric, cryospheric, biospheric, hydrological, oceanographic data;
- Climate Metadata;
- WMO Standard Climate Products;
- Data derived from both in-situ and remotely sensed observations;
- The output generated by numerical models;
- Other related data such as proxy data (tree rings, ice cores etc.).

Climate data management also has challenges that are not currently found with many other types of data, for example:

- Each type of data is considered to be time series data;
- It is becoming increasingly important to understand the provenance of the data;
- Climate data is considered as *Big Data*, Wikipedia-1 (2022) characterised by large volumes of data in a variety of data formats that present challenges for effective data management.

To illustrate the inter-relationship between observations, derived data and the necessary concept of data provenance, which is one of the climate data management challenges, an instructive use case is available under World Wide Web Consortium (W3C) at this address:

https://www.w3.org/2021/sdw/UseCases/SDWUseCasesAndRequirements.html#ProvenanceOfClimateData.

This provides context to the Provenance and Lineage component, later on in this chapter, and reminds users that provenance needs to be traceable from published papers back through climate models and derived data sets, right back to the initial raw observations.

4.1 Observations data

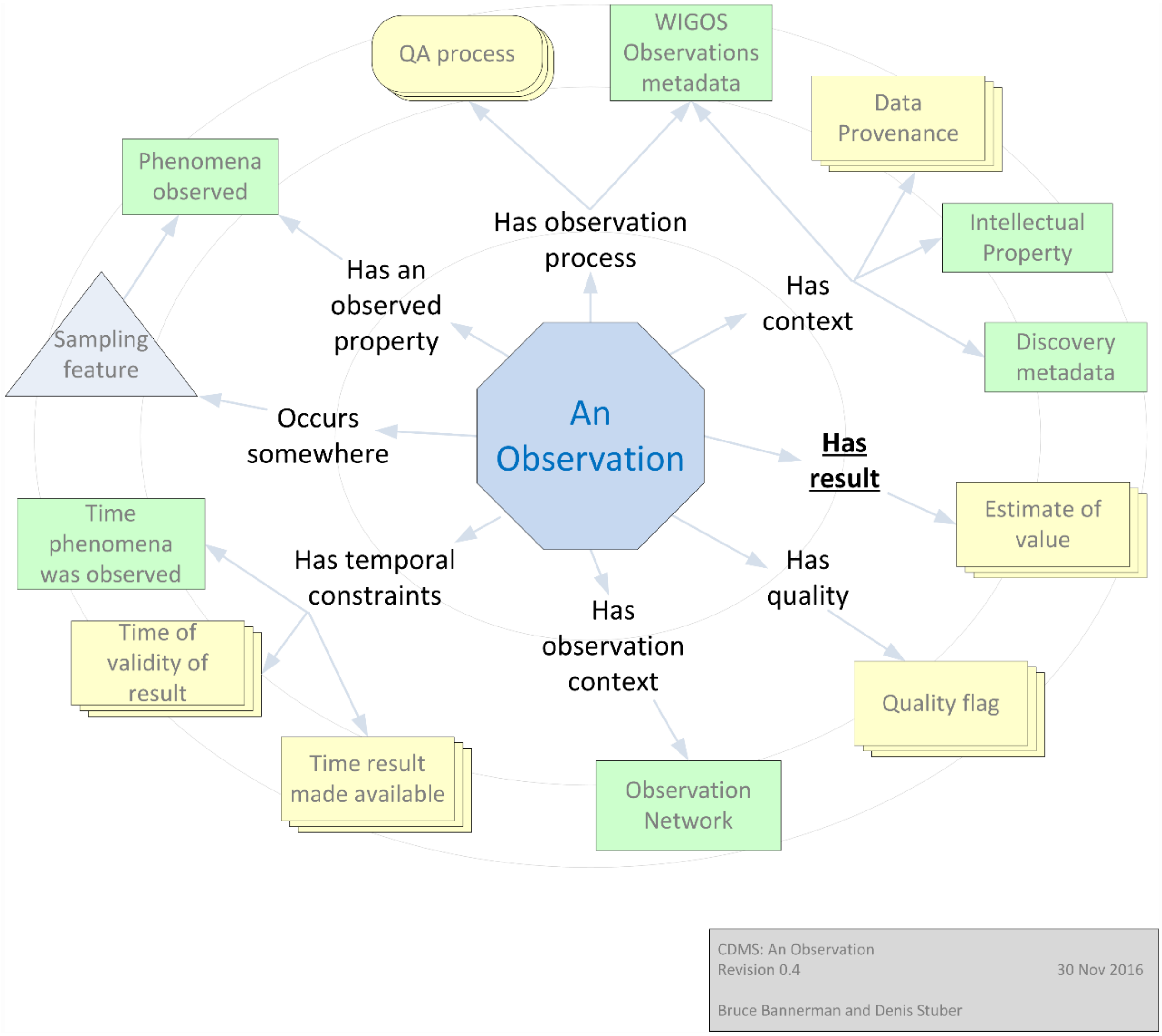


Figure 4.3 Observation concept

Observations data is the foundation of quality climate data. It represents the point of truth describing what phenomena was observed within the climate system at a specific location at a specific point in time. This data is critical for proving the provenance of upstream information, including a variety of derived climate monitoring data, homogenised datasets and numerical models such as climate reanalyses. Therefore it is essential that any change to the results of an observation through a variety of quality control activities are recorded, allowing traceability back to what was originally observed.” The “An Observation Conceptual Architecture” diagram above is an abstract representation aimed to visually represent the many aspects that comprise a quality observation.

In 2010, the Global Climate Observing System (GCOS) published the Essential Climate variables (ECVs). ECVs are physical, chemical or biological variables or a group of linked variables that critically contributes to the characterization of Earth’ s climate. In 2022, the GCOS identified 54 ECVs indexed by the domains of Atmosphere, Ocean, Cryosphere, Anthroposphere, Biosphere and Hydrosphere, as illustrated by the graphic below:

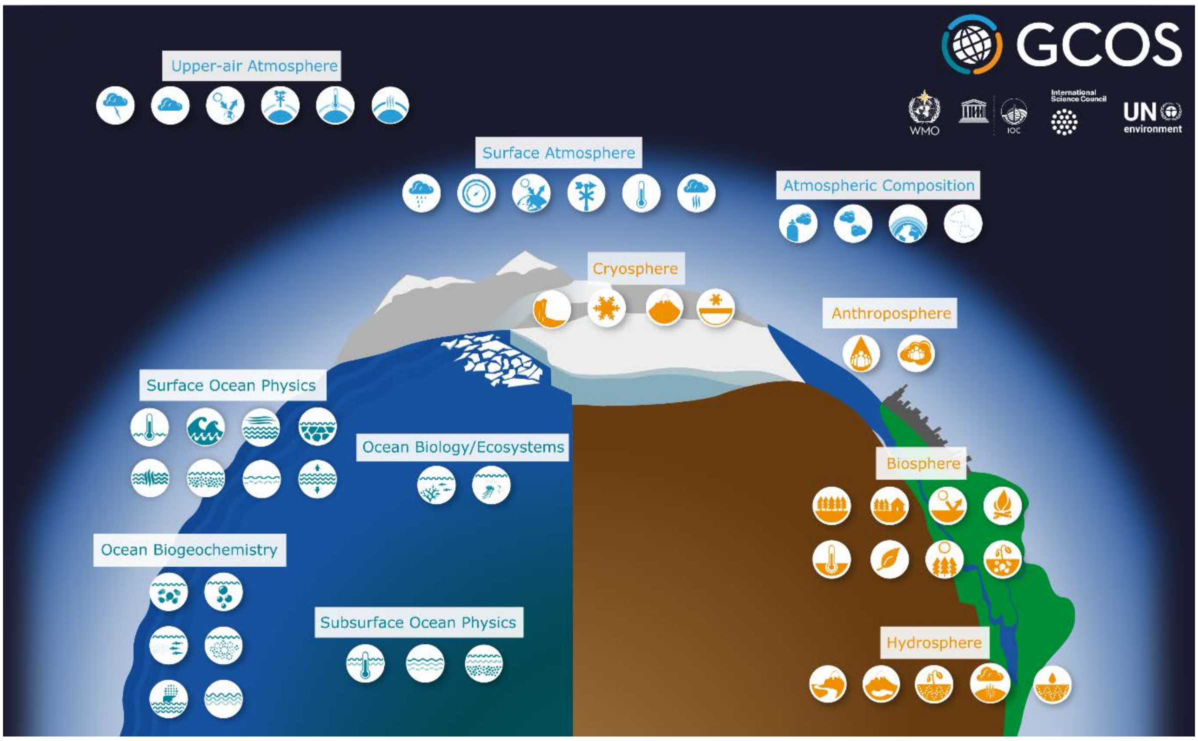


Figure 4.2. Essential Climate Variables from the GCOS
(https://gcos.wmo.int/en/essential-climate-variables, view the 09/03/2023)

For more information see the 2022 GCOS ECVs Requirements (GCOS-245). It is worth noting that there are other types of observations not listed here that are very relevant to climate services, including visibility, evaporation, meteorological phenomena, etc. In the case of oceanographic observations, the Guide to Marine Meteorological Services (WMO-No 471, 2018) lists a wide range of observations beyond the GCOS ECVs, including such variables as icebergs, frozen sea spray, tsunamis, etc.

The GCOS gives requirements on each climate variable e.g. its frequency, resolution, uncertainty, stability…

In parallel, the Observing Systems Capability Analysis and Review Tool of WMO, so called OSCAR, is the official repository of requirements for observation of physical variables in support of WMO Programmes and co-sponsored programmes. WMO is maintaining a database accessible through its web site (https://space.oscar.wmo.int/observingrequirements).

It is essential that CDMS have the ability to monitor data from multiple instruments for each station in an NMHS network, including multiple temperature and precipitation sensors in the case of climate monitoring stations, as per the GCOS

Surface Reference Network. For the latter, the recommended extended requirement on measurement redundancy for temperature is to employ three sensors for added confidence and robustness, as is done, for temperature, for instance, in the US Climate Reference Network (https://www.ncei.noaa.gov/access/crn/instruments.html). However, this level of requirement is surely difficult to achieve for many NMHSs.

This is to ensure that stations:

- Follow the GCOS Monitoring Principles and more specifically the 2 first principles, which are: (1) the impact of new systems or changes to existing systems should be assessed prior to implementation; (2) a suitable period of overlap for new and old observing systems is required. See: https://gcos.wmo.int/en/essential-climate-variables/about/gcos-monitoring-principles
- Be aligned with (1)WCDMP-No. 62 (WMO-TD No.1378, 2007) on "Guidelines for managing changes in climate observation programmes", see the paragraph on "Parallel observations, the preferred approach"; (2) WMO-No. 8 , Guide to Instruments and Methods of Observation, 2021 edition, Volume I – Measurements of Meteorological Variables (paragraph 1.3.4 "Changes of instrumentation and homogeneity" and also Volume III – Observing Systems (paragraph 1.2.3 Instrument"); (3) WMO-No. 1202, 2017, "Challenges in the transition to Automatic Meteorological Observing Networks for long-term climate records"
- Capture information from different co-located instruments, such as a manual raingauge and a tipping bucket raingauge. Such complementarity of instrumentation reduces the risk of gaps in the climate record, and may assist with operations such as disaggregation

Hereafter are listed the different GCOS domains. The required statement, means that the CDMS must be able to accept and handle all kinds of variables from those different categories, noting that in some cases the data may be provided by a third party. It does not mean that the CDMS shall include initially all the variables, but the system should have the capacity to easily incorporate new variables.

4.1.1 **Atmosphere**	Surface and Upper-air atmosphere such as precipitation, pressure, radiation, wind, temperature, lightning, clouds, aerosol, greenhouse gases…	Required
4.1.2 **Ocean**	Surface and subsurface ocean physics, biology, ecosystem, biochemistry, …	Required

4.1.3 **Cryosphere**	Snow, glacier, permafrost…	Required
4.1.4 **Anthroposphere**	Part of the environment that is made or modified by humans such as water use and anthropogenic greenhouse gas fluxes	Required
4.1.5 **Biosphere**	Landcover, above ground biomass, albedo, fire, land surface temperature…	Required
4.1.6 **Hydrosphere**	River discharge, soil moisture, evaporation from land, …	Required

4.2 Logical data models

In order to share climate data between organizations or make meaningful comparisons between information from datasets provided by different publishers, it is highly recommended that climate data conform to a logical data model.

A Logical Data Model (LDM) describes the content and structure of information resources at an abstract level, i.e in a way that is not system-specific.

For a CDMS, a Logical Data Model establishes structure and rules about all elements on data and data management, and describes the relationships among them.

A LDM may be though of as a means to facilitate the consistent exchange of data between disparate system, where differences exist between their data structure. It *is independent of* technology and independent of a specific database implementation. This means that:

- several CDMSs that comply to a specific LDM could come up with different technical solutions and different interfaces;
- most importantly all CDMSs that comply to a specific LDM "talk the same language" and are able to share and store information without ambiguity. That is the basis of interoperability between systems and enables the possibility to automate machine-to-machine exchanges;

While a single LDM covering all types of Climate Data may be desirable, it would be very complex and need to cover a widely disparate amount of concepts from observations and observations metadata through derived data, spatial representations, through to a variety of numeric models. However such a single LDM would be very complex. At present work is underway within WMO on LDMs that are pertinent to climate observations and metadata.The following introduces the components needed to develop WMO logical data models pertinent to climate data. The first step is to gather information on the vocabularies used in this domain (4.2.1), then to present a general view on the existing standards (foundation standards; 4.2.2), and presenting finally the extended WMO logical data models in place or in development (4.2.3).

4.2.1 Controlled Vocabularies

Having authoritative definitions for terms and common vocabularies are crucial to building efficient communications within a community. Hereafter are listed initiatives along these lines, both from WMO and standardization organizations.

At its nineteenth Congress (May-June 2023, Doc. 4.2(3)), WMO decided to develop and maintain the WMO Standard Vocabulary which will become the "authoritative, single entry point" for WMO terminology and related definitions. Until such a tool is

in place, the reader can refer to the following documents and applications which are labelled as “Required” or “Recommended” It is worth noting that efficient communication in general, and powerful interoperability of systems, are the result of (1) a common meaning of things shared by a group of people, i.e., the **semantic**, and (2) an agreement on all elements that define the domain of interest with their relationship to each other, the **ontology**. A Semantic technology approach is today implemented in different applications by organisations such as Media Management, Web Search , etc. See: • Web Ontology Language (OWL) from the W3C, OWL - Semantic Web Standards (w3.org) • Semantic technology (Semantic technology - Wikipedia)		
4.2.1.1 **WMO Documents**	Currently, definitions of terms are scattered across different kinds of WMO publications: Technical Regulations, Manuals, Guides, etc. This is a major stumbling block and the absence of a ‘single entry point’ makes it very difficult for communities inside and outside WMO to identify which publications should be prioritized when seeking an “authoritative definition”. The Requirement here is to adhere as closely as possible to standard terminology in WMO documents. The following documents are where most definitions could be found: • *Technical Regulations* (WMO-No. 49, Vol I) • *Manual on Marine Meteorological Services* (WMO-No 558) • *Manual on WIGOS* (WMO-No. 1160); • *Manual on High-Quality Data Management Framework for Climate* (WMO-No. 1238) • *Guide to Instruments and Methods of Observation* (WMO-No. 8) • *Guide to WIGOS* (WMO-No. 1165) • *WIGOS Metadata Standard* (WMO-No. 1192)	Required

	These references may be found on the WMO Library site.	
4.2.1.2 **WMO Codes Registry**	WMO Codes Registry is a tool that publishes coded structures content from the official WMO Manuals in a web accessible form. This initiative, at first developed for Aviation data exchange, is also now used for the WIGOS Metadata Representation, and for BUFR and GRIB codes. It is accessible at: http://codes.wmo.int/	Recommended
4.2.1.3 **UNTERM**	After the International Meteorological Vocabulary (WMO-No. 182) with the International Glossary of Hydrology (WMO-No. 385), WMO developed MeteoTerm the WMO terminology database. MeteoTerm is no longer supported and has been integrated into the United Nations Terminology Database UNTERM that is accessible at UNTERM.	Recommended
4.2.1.4 **ISO OGC W3C**	Standards come mainly from the International Organization for Standardization (ISO), the Open Geospatial Consortium (OGC) and the World Wide Web Consortium (W3C). It is recommended to follow the authoritative definition terms of those organizations.	Recommended

4.2.2 Foundation standards

This component represents technology that provides rules and a standardized approach for modelling observations data, regardless of the domain.

4.2.2.1 **Observations,**	The Observations, measurements and samples (OMS) Standard is at the same time an OGC and an ISO standard (ISO 19156). OMS is one of the core standards in the OGC Sensor Web Enablement suite,	Required

Measurements and Samples	providing the response model for Sensor Observation Service (SOS). OMS helps to share easily measurement and observation events occurring in different kinds of sensors from thermometers inside an industrial process to satellites taking images of the Earth from space. In essence, the *Observations, Measurements and Samples* standard treats an observation as an event at a given point in time whose result is an estimate of the value of some property of a feature of interest, obtained using a known observation procedure. This standard is being widely adopted as the framework for a number of logical data models related to observations data, such as water data (WaterML) and the Meteorological Information Exchange Model of the International Civil Aviation Organization (ICAO) the IWXXM. It also underpins current works on the WMO logical data model called the Climate Data Model (see below). For more information, see: ISO/DIS 19156, *Geographic information – Observations and measurements* OGC Abstract Specification: Geographic information – Observations, Measurements and Samples	
4.2.2.2 **Coverage**	Based on OGC and ISO standards, the Coverage model establishes a concrete and interoperable coverage structure A coverage structure, depicting spatial variability, is commonly used in meteorology with gridded data to represent Numerical	Recommended

Weather Prediction and Projection model outputs

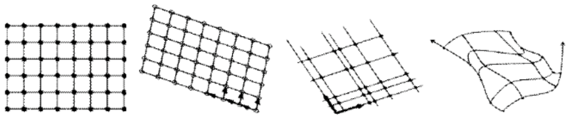


Figure 4.4 Gridded Coverage Types

Coverages may be more than just gridded data. The description of a coverage from the OGC Coverage Implementation Schema: “Coverages represent homogeneous collections of values located in space/time, such as spatio-temporal sensor, image, simulation, and statistics data. Common examples include 1-D timeseries, 2-D imagery, 3-D x/y/t image timeseries and x/y/z geophysical voxel models, as well as 4-D x/y/z/t climate and ocean data. Generally, coverages encompass multi-dimensional, regular and irregular grids, point clouds, and general meshes.”.

Coverages can be encoded in any suitable format (such as GML, JSON, GeoTIFF, GRIB or NetCDF) and can be partitioned, e.g., for a time-interleaved representation. Coverages are independent from service definitions and, therefore, can be accessed through a variety of OGC services types, such as the Web Coverage Service (WCS) Standard.

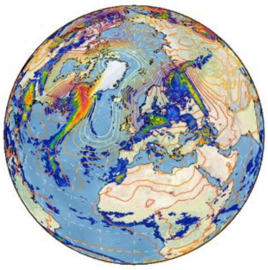
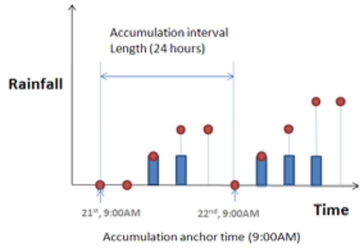


Figure 4.5 Numerical Weather Prediction model output

	Contrary to an image (or raster) such a structure allows the exchange of all the data which compose it and thus gives the possibility of calculations of all natures. Some references: https://portal.ogc.org/files/?artifact_id=19820 https://portal.ogc.org/files/?artifact_id=7467 https://docs.ogc.org/is/09-146r8/09-146r8.html	
4.2.2.3 Timeseries	This OGC standard defines the semantics of observational data as timeseries for data exchange. This allows observational information systems to communicate timeseries data between systems and to end users. This standard has been developed from work initially undertaken within OGC WaterML (see below). Work has been done to remove the hydrology specific aspects of this work to produce a domain-neutral model for the representation and exchange of timeseries data.  *Figure 4.6 Example: Accumulated daily rainfall totals from 9am to 9am.*	Recommended
4.2.2.4 Provenance	"Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness" (W3C Working Group Note).	Recommended

It enables the interoperable interchange of provenance information in heterogeneous environments.

Lots of application areas need provenance and especially the science domain in order to provide background on how the results were obtained, including any changes made. Refer to the 4.3.3 Data Provenance component later in this chapter that lists the important information that should be kept to explain any modifications on the data.

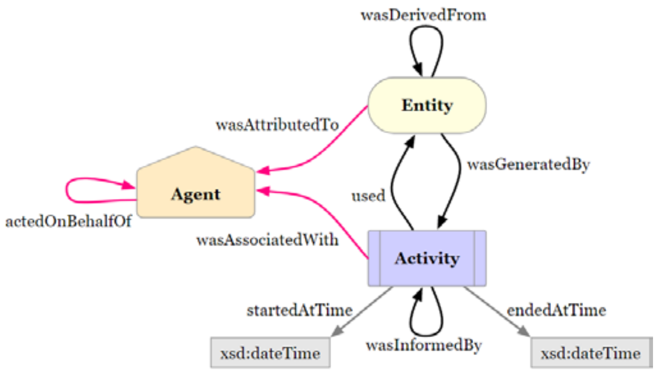


Figure 4.7 W3C the PROV Ontology

This data model will help the long term sustainability of provenance data, since it relies on interoperable formats and refers to a generic data model. Its potential to be adapted to a particular domain and provenance use case (by including domain vocabularies and by specialising its classes) is very important and requires investment.

Early adoptions of PROV in Climate Services can be referenced with the following:

Metaclip:
https://www.sciencedirect.com/science/article/abs/pii/S1364815218305036

Roocs
https://rook-wps.readthedocs.io/en/latest/prov.html

	ESMVal https://docs.esmvaltool.org/projects/ESMValCore/en/latest/interfaces.html SWIRRL / Climate4Impact workspaces https://direct.mit.edu/dint/article/4/2/243/109846/SWIRRL-Managing-Provenance-aware-and-Reproducible SPRINGER Link https://link.springer.com/chapter/10.1007/978-3-030-52829-4_12 It is worth mentionning here the provenance template work by L Moerau, since it helps adopting the model also as a tool to discuss and drive the provenance requirements, also identifying reusable scenarios. 10.1109/TSE.2017.2659745.	
4.2.2.5 **Discovery Metadata**	The International Organization for Standardization, or ISO, created the ISO 19115 geospatial metadata standard. Using this standard allows users to discover easily, and in a standard way, information about geospatial data which is generally the case for meteorological data. ISO 19115 contains hundreds of fields which could be broken down into categories of information such as Identification, Distribution, Quality, Lineage and Metadata. Using this standard ensures that the process of metadata discovery is interoperable, and conforms to a specific set of rules and standards.	Required

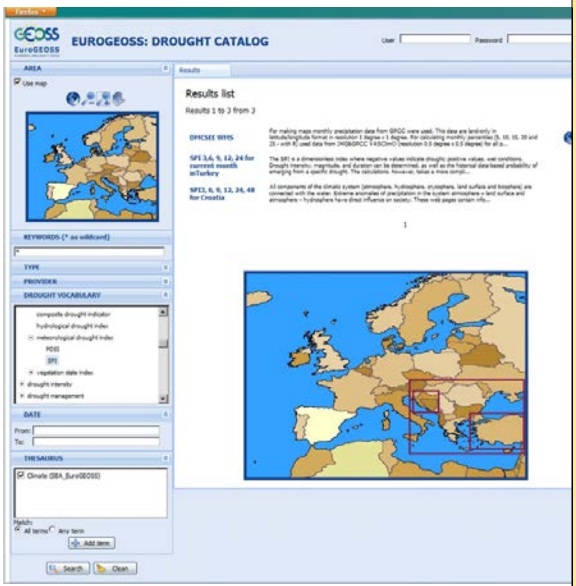


Figure 4.8 Graphical User Interface Web Application for Searching Metadata

ISO 19115 is the basis for WMO Information System version 1 (WIS 1).

When it comes to WIS 2, it is the OGC API - Records that offers the capability to create, modify and query metadata on the web. The specifications of the OGC API Records provide a record metadata model in support of resource discovery.

4.2.3 WMO logical data models

It is always difficult in data modelling to precisely define which logical data model (LDM) is required, recommended or optional. However, it is strongly recommended here that the *WIGOS Metadata Standard (WMO No 1192)* be a required onefor CDMSs developed for WMO Members. If there is general agreement that the WIGOS metadata standard shall be “Required”, then this forms the basis for extended versions of LDMs which should be chosen with care according to the system to be developed, the expected benefits, and the latest advice from WMO that may evolve in this technological domain.

4.2.3.1 **WIGOS Metadata**	According to the WIGOS Metadata Standard (WMDS, WMO-No 1192), “The WMO Integrated Global Observing System observations consist of an exceedingly wide	Required

	range of data, from manual observations to complex combinations of satellite hyper-spectral frequency bands, measured in situ or remotely, from single dimension to multiple dimensions, and those involving processing. A user should be able to use the WIGOS metadata to identify the conditions under which the observation, or measurement, was made, and any aspects that may affect its use or understanding, that is, to determine whether the observations are fit for the purpose”. The WIGOS Metadata Representation (WMDR) implements concepts of the WMDS. The “CDMS Specifications” document includes all of the WIGOS Metadata Standard (See 4.3.1 Observation Metadata). For additional information, refer to: • *WIGOS Metadata Standard*, WMO-No. 1192, 2019 • *WIGOS Metadata Representation*, Manual on Codes, Volume I.3, part D – Representation derived from data models WMO-No. 306	
4.2.3.2 **Climate Data**	In 2022, WMO started to develop a generic Climate Data Model for the representation, exchange and archival of data within Climate Data Management Systems (CDMSs), in line with the WIGOS Metadata Representation. The Climate Data Model will be based on OMS and it is proposed that this be adopted by the WMO community. At the time of the release of the second edition of WMO-No. 1131, the generic Climate Data Model is still in development. When finalized the Climate Data Model would surely become a required component. See: https://community.wmo.int/en/governance	in development

	/commission-membership/commission-observation-infrastructure-and-information-systems-infcom/standing-committee-information-management-and-technology-sc-imt/expert-team-metadata-standards-et-metadata/task-team-climate-data-model-tt-cdm	
4.2.3.3 WMO Core Metadata Profile (WCMP)	The primary purpose of WCMP is to describe and discover datasets and services within the WMO Information System (WIS) See the Manual on WIS (WMO-No. 1060) and the Guide to WIS (WMO-No. 1061)	Required
4.2.3.4 WaterML	Based on OMS and developed originally for the hydrometeorology domain, WaterML offers a solution to easily exchange time series on water data. See [WaterML - Wikipedia]	Recommended
4.2.3.5 METCE	METCE, « Modèle pour l'Echange des informations sur le Temps le Climat et l'Eau », or in English “"METeorological Community Exchange model” provides an information framework for modelling WMO codes and other data representations. It allows among other things the ICAO Meteorological Information Exchange Model (IWXXM) for representing observations, forecast and reports, including: • routine meteorological reports (METAR) • aerodrome special meteorological reports (SPECI) • aerodrome forecast (TAF) reports	Optional

	• SIGMET information • AIRMET information • Tropical Cyclone Advisory and Volcanic Ash Advisory See Manual on Codes, Volume I.3, part D – Representation derived from data models (WMO-No. 306).	

4.2.4 Climate database

This subsection refers to the database(s) used to store and manage a range of time-series data, including: climate observations, climate metadata (observations, discovery, and data provenance), spatial information, derived data, and related data required for effective data management and interpretation.

More advanced CDMSs may manage the data in a series of related databases rather than in a single database.

It is recommended that the climate database provide support for the following functionalities, classified by priority:

Required

- Managing core observations described in the Guide to Climatological Practices (WMO-No. 100)
- Managing observation metadata (such as station and instrumental metadata) and integrating, i.e explicitly linking them, with observations data
- Managing metadata related to data provenance. This entails ensuring that each change to an observation is recorded for future recovery. This involves recording the details of why a particular change was made, which includes: tracing the product lineage to the data source. For example, what observations and gridded data were used to underpin the analysis released in peer-reviewed paper X?
- Handling observations from multiple sensors per station, per phenomenon, and recording the source of each observation
- Managing multiple tiers of data quality, from raw records through quality controlled and homogenized data
- Managing spatial and time-series data
- Covering the GCOS ECVs and the variables listed by in the OSCAR/Requirements database (https://space.oscar.wmo.int/variables)

Recommended

Using a robust data model that takes into account the requirements of open spatial standards, particularly the ISO 19156:2023 *Geographic information – Observations and measurements* standard, and the WMO standards such as the WIGOS MetaData Standard.

- Managing third-party and crowdsourced data
- Managing intellectual property rights related to data
- Enabling point-in-time recovery. For example, what data were present in the database for station X at time T?
- Storing a range of document formats and types, such as:
- Photographs of observation stations and instruments, meteorological phenomena, etc.
- Scanned paper observation forms
- Scanned microfiche/microfilm
- Relevant observations metadata documents, such as instrument calibration reports
- Technical and procedural manuals
- Site location plans and sections
- Videos and other multimedia formats

Optional

- Handling data uncertainty (for more information, see Wikipedia articles on uncertain data and uncertainty)
- Managing multidimensional time-series gridded data and possibly numerical model data fields
- Providing support for the information management concepts of semantics and linked data

4.2.4.1 Data dictionary	This component represents the data dictionary, which describes the database structure, data model and other elements used by the climate database. In another words it is where the collection of tables of the database(s) are described (meaning, relationships to other data, origin, usage and format).	Required

4.3 Climate metadata

Climate Metadata is defined in this document as referring to the suite of supporting data that is required to effectively manage climate data and assess its fitness for purpose.

This version of WMO No 1131 has been updated to incorporate advances in, and new requirements for, standards (notably the WIGOS Metadata Standard, WMDS) in relation to all observational data, but specifically those associated with climate data.

Climate Metadata comprises the following components:

Observations Metadata

Time series data about how, when and where meteorological observations were made, and the conditions that they were made under.

Discovery Metadata

Intended to facilitate the discovery and assessment of a dataset to determine if it is fit for re-use for a purpose that may be different from that for which it was originally created.

Data Provenance Metadata

Information relevant to climate data that details what has been done to the data (quality control, homogenisation etc), and thereby facilitates end users, including data managers, scientists and the general public, to develop trust in the integrity of the climate data.

Each of these components of Climate Metadata is expanded upon in this section of the CDMS Specifications.

4.3.1 Observations metadata

This component supports access to, and management ofthe observed variable, the conditions under which it was observed, how it was measured, and how the data have been processed, in order to provide users with confidence that the data are appropriate for their application. It is used to support a range of activities that allow climate professionals to understand the fitness for purpose of specific data and in many cases improve the quality of climate observations data.

WMO developed the WMO Metadata Standard, WMDS (WMO-No 1192), to provide a common language and a common set of expectations for Observations Metadata both on the surface and in space. The WMDS makes it easier to create, share and integrate data by ensuring that the data are represented and interpreted in an agreed

systematic way. Such a standard facilitate interoperability between the different domains of meteorology: Atmosphere, Ocean, Cryosphere, Anthroposphere, Biosphere and Hydrosphere. The focus of this component is is on surface observational data, but in future it should later be extended to other types of observational climate data. The WIGOS Metadata Standard (WMDS) consists of ten categories, each category consisting of one or more metadata elements. Added to those categories, there must be bridges between WMO's two global databases, the OSCAR/Surface & Space and the OSCAR/Requirements.		
4.3.1.1 **Observed variable**	Gives the basic characteristics of the variable (temperature, precipitation, sea-level, streamflow) being measured, and the resulting datasets. It includes an element describing the spatial representativeness of the observation as well as the biogeophysical compartment the observation describes.	Required
4.3.1.2 **Purpose of observation**	Represents the main application area(s) of the observation and the observing programme(s) and networks the observation is affiliated with.	Required
4.3.1.3 **Station / platform**	Represents the observing facility, including fixed station, moving equipment or remote-sensing platform, at which the observation is made.	Required
4.3.1.4 **Environment**	Describes the geographical environment of the site where the observation is made. It also provides an unstructured element for additional meta-information that is considered relevant for adequate use of the observations and that is not captured anywhere else in this standard.	Required

4.3.1.5 Instruments and methods of observation	Describes the method of observation and the characteristics of the instrument(s) used to make the observation. If multiple instruments are used to generate the observation, then this category should be repeated.	Required
4.3.1.6 Sampling	Represents how sampling and/or analysis are used to derive the reported observation or how a specimen is collected. An example might be how many one second observations are required to define a one minute observation from an AWS.	Required
4.3.1.7 Data processing and reporting	Represents how raw data are transferred into the observed variables and reported to the users.	Required
4.3.1.8 Data quality	Represents the data quality and traceability of the observation.	Required
4.3.1.9 Ownership and data policy	Represents who is responsible for the observation and owns it.	Required
4.3.1.10 Contact	Represents where information about the observation or dataset can be obtained.	Required

1

4.3.2 Dataset discovery metadata

This component refers to the processes, software and governance arrangements that ensure that discovery metadata are captured, managed and maintained.

Discovery metadata are intended to facilitate the discovery and assessment of a dataset to determine if it is fit for reuse for a purpose that may be different to the purpose for which it was originally created.

Discovery metadata may also be known as WIS metadata. **They are not the same as the observations metadata described above.**

4.3.2.1 Dataset identifier	This component represents a unique identifier used to identify the dataset.	Required
4.3.2.2 Dataset overview	This component gives a text overview of a dataset. This may include a description of the dataset (such as an abstract), the intended use of the dataset, its lineage and status.	Required
4.3.2.3 Dataset data quality	This component represents a general assessment of the quality of a dataset.	Required
4.3.2.4 Distribution	This component covers information about the distributor of, and options for obtaining, a dataset.	Required
4.3.2.5 Access constraints	This component provides information on the restrictions in place for a dataset.	Required
4.3.2.6 Dataset maintenance	This component provides information on the scope and frequency of updates and maintenance conducted on a dataset.	Required

4.3.2.7 Spatial representation	This component covers information on the mechanisms used to represent spatial information within a dataset.	Required
4.3.2.8 Reference systems	This component gives information on the reference systems used by a dataset. These include a horizontal spatial reference system, vertical spatial reference system and temporal reference system.	Required
4.3.2.9 Stewardship Maturity Matrix	In 2023 WMO published the Guidance booklet on Maturity Matrix for Climate Data (2023, WMO-No.1328). This component measures the level of capability of stewardship practices applied to datasets. It generally defines measurable, level-progressive practices of key elements of stewardship such as preservation, accessibility, metadata, quality control, and transparency/traceability, rating each element on a level scale from not managed to optimally managed. Such elements should guide users in selecting the dataset that best meets their requirements, and should also be a guide to CDMS owners on where improvements to stewardship practices are required. Such stewardship assessments could become part of the discovery metadata for datasets.	Recommended

4.3.3 Data provenance and Lineage

This component refers to the processes, documentation and governance arrangements that record and manage information relevant to climate data and enable end-users (including data providers, scientists and the general public, to develop trust in the integrity of the climate data).

Data provenance establishes the authenticity and reliability of data by allowing an end-user to trace back to its source and understand the transformations it underwent. It also helps the user to identify what version of the data was available at any given time.

The need for this new type of climate metadata has become more evident following consistent attacks on the credibility of climate data. One notable example is the so-called Climategate incident and subsequent inquiries.

Therefore, it is important for NMHSs to establish the reliability of their climate data and processes and to ensure that these data are subsequently seen as the authoritative source that can be used for global climate studies. Also, if changes have been made, it is important to state defensible reasons for changes, to avoid any perception that changes made were to support non-scientific ideologies.

While the concept of data provenance has been relatively nebulous within the information management domain for many years, there has been a significant amount of work on the concept within the World Wide Web Consortium (W3C) for over a decade, particularly with regard to the development of the PROV standard.

The W3C defines provenance as:

"A record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing. In particular, the provenance of information is crucial in deciding whether information is to be trusted, how it should be integrated with other diverse information sources, and how to give credit to its originators when reusing it. In an open and inclusive environment such as the Web, where users find information that is often contradictory or questionable, provenance can help those users to make trust judgements." (See PROV-DM: The PROV Data Model (w3.org))

While this work is still relatively new, it is showing significant potential for use within the climate domain.

Maintaining data and datasets with high levels of data provenance metadata may be challenging. Therefore, priority should be given to data of high importance and visibility such as high-quality climate monitoring datasets. It is anticipated that guidance will be required to suggest what data should be maintained with what level of data provenance metadata. This guidance should be included as a policy within NMHSs' climate data management governance frameworks.

For more information, see:

- Overview of the PROV family of documents
- PROV data model specification
- Lineage information from ISO 19115

At the begining of chapter 4, a W3C use case is mentioned that gives the importance of Provenance from published papers to raw data. Let's take another concrete (but simplified) example of missing data from a weather station.

Following a telecommunications breakdown at an Automatic Weather Station (AWS), data was not transmitted to the headquarters of a National Meteorological and Hydrological Service (NMHS) for a period of 10 days, from 24 August to 2 September. These non-transmitted data, minute or hourly data, are considered as missing by the system, which in most cases is software called a Climate Data Management System (CDMS) monitored by the NMHS.

According to its contractual obligations the NMHS provides clients with daily and monthly evapotranspiration data from this specific AWS. The daily data is provided to the customer every day and the monthly data before the second day of the following month.

Therefore, in order to fulfil its obligations, the NMHS uses appropriate techniques to estimate the missing values from 24 August to the 2nd of September. It must estimate the values for 10 daily evapotranspiration values and the monthly evapotranspiration value for August using a reliable and documented scientific procedure. These data, recognized as “estimated” by the CDMS, are sent to the clients with the appropriate metadata information.

On 3 September, the NMHS finally received the minute or hourly raw data (24 August-2 September) from the relevant AWS. The NMHS is now able to use this raw data to perform the following tasks:

1. Import the retrieved raw data into the CDMS that will supersede the estimated values;
2. Create/modify the provenance and lineage information for this raw data;
3. Start the data quality control chain for these raw data;
4. Compute the daily and monthly evapotranspiration for 10 days and 1 month;
5. Create/modify the provenance and lineage information for these 11 values;
6. Notify users that the daily and monthly evapotranspiration values for this AWS have been changed for the relevant dates.

We can see that, even in a simplified way, this case study of missing and then retrieved data involves many actions that need to be automated. We can also see that it may seem easy to gather information that can answer the following questions, for example as part of the description of a new version of the dataset(s):

- ✓ What has changed? (e.g. what part of the data)
- ✓ Why was it changed
- ✓ Who/what changed it?
- ✓ On whose behalf did they act?
- ✓ When was it changed?
- ✓ Who was responsible?
- ✓ What was it derived from?

However, it will be more difficult to capture details of the underlying process that updated the gap, as technical details concerning points 1 and 4 need to be recorded as lineage information i.e. the process of recording details of data transformations, and will therefore require automated mechanisms to do so. In this way, the following questions should be answered:

- ✓ How was the raw data processed?
- ✓ How was the daily and monthly data processed?

Combining the information collected to answer the above questions, it will be possible to fill in change-log records to answer the following questions:

- ✓ How was it changed?
- ✓ What was done to change it?

4.3.3.1 **What was changed?**	This component refers to the processes, software and governance arrangements that ensure that any change to climate data is recorded.	Required
4.3.3.2 **When was it changed?**	This component covers the processes, software and governance arrangements that ensure that the time of the change is recorded.	Required
4.3.3.3 **What was it derived from?**	This component deals with the processes, software and governance arrangements that ensure that a dataset’s lineage is understood. In other words, where did the data come from? This component is also required in the section on data discovery (8.2).	Required

4.3.3.4 What was done to change it?	This component refers to the processes, software and governance arrangements that ensure a clear explanation of any ad hoc modifications to a climate record. This includes: • What was changed • When the change was made • Details describing what was done	Required
4.3.3.5 How / why was it changed?	This component refers to the processes, software and governance arrangements that ensure that the rationale behind a modification to a climate record is clearly understood. This includes: • How the change was made? • Why it was made?	Required
4.3.3.6 Who / what changed it?	This component involves the processes, software and governance arrangements that ensure a clear understanding of the agent that affected the change.	Required
4.3.3.7 Who did they act on behalf of?	This component refers to the processes, software and governance arrangements that ensure that the person or role who requested the change, or authorized the software, is identified.	Required
4.3.3.8 Who was responsible?	This component refers to the processes, software and governance arrangements that ensure that the person, role or policy that provided the authority to make the change is identified.	Required

1

4.4 WMO standard products

4.4.1 Observation data products

This subsection outlines the types of data products, and in particular climate-related products and reports, that NMHSs have committed to generate and provide to WMO. The architecture of the WMO Information System (WIS) provides for the submission and exchange of such products

4.4.1.1 Routine messages	This component represents data computed from routine observationsfor use in WMO products. There are numerous such messages and reports as outlined in the Manual on Codes (WMO-No 306). For example, the SYNOP, CLIMAT and DAYCLI messages. For more information, see: • *Manual on Codes* (WMO-No. 306), 2022 edition.	Required
4.4.1.2 Climatological normals	This component covers monthly and annual climatological standard normals. Climatological standard normals comprises the following: A fixed reference period (1961–1990) for long-term climate variability and change assessment. This is to be adopted as a recommended stable WMO reference period, until such time as there is a compelling scientific case for changing it. A varying standard 30-year period updated every 10 years to be employed for all climate services. The current period is 1991-2020. Whichever baseline is used, 1961-1990 or 1991-2020 or others should always be	Required

	specified. Enough historical and geographical context should be provided for the user of the information to understand the significance of the information that is being presented. It is recommended that when baselines are updated, previously calculated normals are retained (Refer to component 4.3.3 on Provenance). For more information, see: • WMO Guidelines on the calculation of climate normals (WMO-No. 1203) • Guide to Climatological Practices (WMO-No. 100), section 4.8 Normals • Manual on Codes (WMO-No. 306) • 1961–1990 Global Climate Normals (CLINO) (WMO-No. 847) • The Role of Climatological Normals in a Changing Climate (WMO/TD-No. 1377), WCDMP-61 • WMO-No. 1307, WMO SERCOM-2 Decision 5.5(2)/1 on the “Guidance on the use of climatological standards normals and other baselines in monitoring the state of the climate” • WMO–No. 1238 (Manual on the High Quality Data Management Framework for Climate).	
4.4.1.3 CLIMAT	This component concerns CLIMAT messages that report monthly climatological data assembled at land-based meteorological surface observation sites For more explanation, see: • Handbook on CLIMAT and CLIMAT TEMP reporting (WMO/TD-No. 1188) • Practical Help for Compiling CLIMAT Reports (WMO/TD-No. 1477), GCOS-127 • Manual on Codes (WMO-No. 306), Volume I.2, Annex II to the WMO Technical Regulations: Part B – Binary	Required

	Codes, Part C – Common Features to Binary and Alphanumeric Codes	
4.4.1.4 DAYCLI	This component concerns the newly-required DAYCLI reports, which provide daily values of temperature and precipitation (rain and snow) submitted monthly. The CDMS is required to generate these messages together with the following information: the siting classification, the measurement quality classification, temperature computation method, definitions of climatological day and data quality flags assigned following QC of each daily value. A DAYCLI guidelines is currently under review	Required
4.4.1.5 World Weather Records	This component covers the World Weather Records reports For more details, see: • Guidelines for the submission of World Weather Records 2011+ 2021 edition	Required
4.4.1.6 Ocean, Agricultural, Hydro, Aeronautical etc.	This component covers other WMO standard products, as may be required. Examples include: **Hydrological domain**: reservoir estimation, flood management, irrigation and drainage, energy, water quality, etc. (see WMO-No. 168, Guide to Hydrological Practices). Quality controlled hydrological data are reported to hydrological data centers (e.g. the Global Runoff Data Centre, GRDC) to produce statistical products such as mean monthly data, and seasonal summaries of primary hydrological values. **Agricultural domain**: Crop response, vegetation indices, etc.	Recommended

	Ocean and Marine domains: a range of standard data products through the WMO Marine Climate Data System (MCDS): thermocline, sea level, salinity, etc.	

4.4.2 Climate Monitoring

An international effort has been made to develop, calculate and analyse a suite of climate indices so that individuals, countries and regions can calculate the indices in exactly the same way such that their analyses will fit seamlessly into the global picture. These indices describe the state of the climate system and assist with monitoring climate. They have been developed specifically to assess changing frequency, intensity and duration of climate extremes, including measures of extrema or the number of occurrences above or below a specified threshold.

Similar indices exist for hydrological and marine climate monitoring.

Those indices have been split in two categories: core indices and other approved indices. There are also indices developed by different research groups

4.4.2.1 Core indices	This component represents the core climate change indices as defined in the website of the former Expert Team on Climate Change Detection and Indices (ETCCDI). ETCCDI website These core indices have subsequently been expanded under the auspices of WMO to include more sector-specific indices calculated using the software package Climpact with associated observational data sets available via the Climdex project. The software and data sets will be continually maintained and upgraded; it is advisable to keep abreast of new developments, with reference to both the ET-CCDI and ClimDex Websites (See Climdex, https://www.climdex.org) References: • Climpact https://github.com/ARCCSS-extremes/climpact	Required

	• ClimDex https://www.climdex.org/ National Hydrological Service reporting stations provide river discharge data to the Global Runoff Data Centre (GRDC), and comprise a network of “Climate sensitive stations” used to enable the monitoring of this important Essential Climate Variable on a global basis. Reference: https://www.bafg.de/GRDC/EN/01_GRDC/grdc_node.html;jsessionid=8B54714EBEC926A236F366E90A095907.live11294	
4.4.2.2 Specific indices	This component covers other climate and climate change indices developed by different NMHSs, research groups and other organizations. Examples include drought index, heat and cold waves, Soil Water Index (SWI), Green Vegetation Cover, Leaf Area Index, Surface soil Moisture etc. See WMO-No. 1173 *Handbook of drought indicators and indices*, and the *Guidelines on the definition and characterization of extreme weather and climate events* (WMO–No. 1310) Also, the WMO-No 1287 *Developing the Climate Science Information for Climate Action* and its 2 annexes, Annex I: *Guidance on Methods*, Tools *and Data* and Annex II: *Use cases*.	Recommended
4.4.2.3 National indices and monitoring products	This component represents the indices and other products of the National Climate Monitoring Products, a standardized set of national monitoring products focused on those products that are widely used, enabling broad participation in global climate monitoring. The CDMS should be flexible enough to accommodate national climate monitoring products and other national indices.	Recommended

	For further information on the indices and methods of calculation, see WMO-No 1204. In the hydrological domain, Several National Hydrological Services have defined “reference networks” suitable for detecting climate signals (e.g. Canada, Australia).	

4.5 Derived climate data

4.5.1 Derived observation data

This subsection describes a range of derived observational data products generated from climate observation variables. Rules for data precision, rounding and missing data should be implemented into the CDMS and these are described further in Chapter 3.

4.5.1.1 Normals and averages	This component represents any normals and averages used by NMHSs that are in addition to climatological standard normals. For example, the component should be able to compute: Averages over specified time periods (such as yearly, monthly, 10-day, 5-days, daily, hourly, 6-minute, 5-minute, minute…) Longterm averages for any period (for example 5, 10, 30 or 100 years, seasonal periods or 3-month period…) Close attention should be given to new constraints arising from climate change, e.g. *Estimating daily climatological normals in a changing climate*" (DOI : 10.1007/s00382-018-4584-6) For more information, see: The *Role of Climatological Normals in a Changing Climate* (WMO/TD-No. 1377), WCDMP-61	Required

	Computation of daily and monthly data from hydrological raw data is standard practice at the majority of National Hydrological Services. These data are provided to the Global Runoff Data Centre (GRDC). The calculation of annual characteristics of lowest, mean lowest, mean, mean highest, and highest discharge values based on daily or monthly data is widespread amongst NHSs. Similar calculations are done on the same statistics for consecutive calendar years. Furthermore, seasonal variability of river discharge based on daily or monthly data are also calculated.	
4.5.1.2 **Basic derived variables**	The simplest examples are meteorologically-vital elements such as Mean Sea level pressure, dewpoint, relative humidity, etc. WMO formulae and practices are as described in WMO-No. 8.	Required
4.5.1.3 **Homogenized data**	This component represents high-quality homogenized time-series datasets. Such datasets aim to ensure that the only variability remaining in the time series is that resulting from actual climate variability. See also the subsection on data homogenization (6.1.3).	Recommended
4.5.1.4 **Advanced Calculation**	This component deals with derived data computed from observations for NMHS products. Examples of other types of computed products are summarized below, the computation of these shall be in accordance with the climatology policies in place within the NMHS.	Recommended

	Some examples are: Data generation of sums, averages or extremes, such as generating ten-day data from daily data or daily data from hourly data. The generation of particular data derived from raw data, such as computing the potential evapotranspiration output for agricultural purposes, return periods... The generation of any statistical parameter required for products, such as extreme value analysis. Climate indices such as computed teleconnection indices or also data fusion e.g. surface observation merged with remote sensing acquisition. An example might be rainfall analysis images derived from a combination of gauge measurements and radar observations. NHSs may derive and store advanced hydrological products such as flow duration curves.	
4.5.1.5 Other	This component concerns any other derived observation data product not mentioned above that is required for NMHS purposes.	Optional

4.5.2 Gridded spatial distribution of observations

This subcomponent concerns the capacity to generate or manipulate gridded data according to different techniques such as interpolation and extrapolation.

Some of these techniques are described in the *Guide to Climatological Practices* (WMO-No. 100, Fourth edition), section 3.5.8 Data estimation

These gridded data are spatial data. They are included in this section to show their lineage as a type of derived climate data.

4.5.2.1 Analysed data	This component refers to spatially distributed gridded data that have been derived from observational data as the result of an analytical process. Some examples are: • Singular variables such as: ✓ Normals ✓ Observations for a given day or time ✓ Averages ✓ Percentiles ✓ Cumulative data ✓ Extremes ✓ Homogenized data • Multivariables such as: ✓ The generation of anomalies (difference between the normals data and a specific monthly variable) ✓ More complex data such as potential evapotranspiration	Recommended
4.5.2.2 Other	This component covers any spatially distributed gridded data product not mentioned above that are required for NMHS purposes.	Optional

1

4.5.3 Numerical models

Note: The infrastructure, software and skills required to operate numerical models are undoubtedly beyond the reach of many NMHSs.

As the output from numerical models is of interest to most NMHSs, it is expected that such output will be available via a number of sources, including Regional Climate Centres and WIS.

Therefore, the CDMS should have the ability to work with such data.

4.5.3.1 Climate Predictions and projections	This component refers to the data output from a variety of climate modelling processes. Such data are generally represented by multidimensional array grids. Some examples are: Climate models (such as global climate models) – numerical representations of the climate system based on physical, biological and chemical rules. They vary on timescales ranging from seasonal to centennial. Climate models are often used to produce climate change projections. Downscaled models – derived from climate models but at a much higher resolution to support regional and local analysis. For some purposes Numerical Weather Prediction Grids may be useful, such as for the purposes of Quality Control. See also: • *Guide to Climatological Practices* (WMO-No. 100), section 4.4 Climate Products	Recommended
4.5.3.2 Reanalysis	Designed for climate monitoring and climate studies, reanalyses provide gridded data over a long time period. Reanalyses are created via an unchanging (frozen) data assimilation scheme and model(s) which ingest all available observations, and utilizes best-practice data assimilation techniques and updated datasets. This unchanging framework provides a dynamically consistent estimate of the climate state at each time step. Some available reanalysis products include ERA-5 (1959 to present) from the European Centre for Medium-Range Weather Forecasts, and the Twentieth Century Reanalysis project (1836–2015) from the National Oceanic and Atmospheric Administration, United States.	Recommended

	See: NCAR, Atmospheric Reanalysis: Overview & Comparison Tables at https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables	

4.6 Ancillary data

This section covers a range of data that complements the climate record stored within the CDMS. While much of this will be provided, and often held, by external agencies, a CDMS should be able to work with such data

4.6.1 Spatial

4.6.1.1 Topography	Although this component is labelled topography, it actually refers to a wider set of data. Some examples are: • Typical topographic data such as drainage, relief, cultural and nomenclatural features • Digital elevation models • For hydrological purposes, ancillary data are generated on topography of hydrological gauging stations, such as size of upstream basin, gradient of rivers, and other features	Recommended
4.6.1.2 Emergency management	This component concerns datasets that are useful for supporting emergency management and related warning systems. This component is Optional, because while access to this information is important, it would likely be more efficient to link to other appropriate data sources within a country, such as emergency management authorities etc.	Optional

4.6.1.3 **Agricultural**	This component refers to agricultural information datasets. Some examples are: • Data from the Food and Agriculture Organization of the United Nations that could relate to agriculture, animal production and health, fisheries, forestry, land and water or plant production and protection. • Regional, national or international data from different organizations such as primary industry departments or research centres on agriculture. This component is Optional, because while access to this information is important, it would likely be more efficient to link to other appropriate data sources within a country, such as agricultural departments etc.	Optional
4.6.1.4 **Health**	The component refers to a wide variety of health datasets. Some examples are: • Data from the World Health Organization covering datasets on a very large spectrum. • National or international data from different organizations such as health departments or health research centres. • Epidemiological studies (see Wikipedia article) and so forth. This component is Optional, because while access to this information is important, it would likely be more efficient to link to other appropriate data sources within a country, such as health authorities etc.	Optional

4.6.1.5 **Environmental**	This component refers to environmental datasets. Some examples are: • Data from the United Nations Environment Programme or the United Nations Educational, Scientific and Cultural Organization. • National or international data from different organizations such as environment departments or research centers. This component is Optional, because while access to this information is important, it would likely be more efficient to link to other appropriate data sources within a country, such as environmental departments.	Optional
4.6.1.6 **Phenological data**	Phenological data, i.e. observations of changes in natural systems such as flowering, bird migrations etc, provide very valuable information on the effects of climate change. Such data may be stored within the CDMS or obtained by a suitable link to research organizations, etc.	Recommeded
4.6.1.7 **Administrative**	This component covers administrative data. Some examples are: • Localities and gazetteers • Administrative boundaries • Transportation networks • Land registry	Recommended
4.6.1.8 **Impacts data**	This component concerns a range of spatial data that relate to things impacted by climate. This could include: • Deaths caused by heatwaves, prolonged droughts, floods, cyclones, fires etc.	Recommended

	• Infrastructure and environmental damage caused by a range of events such as floods, bush fires or cyclones. • Changing land use, such as agricultural adaptations due to a changing climate. While access to this information is important, it would likely be more efficient to link to other appropriate data sources within a country.	
4.6.1.9 Other	This component refers to a range of other spatial data that may be relevant to climate.	Optional

4.6.2 Climate documentation

4.6.2.1 Published reports	This component represents the processes and governance arrangements that result in the preparation and release of a wide variety of written reports. Some examples are: • Climate change impact studies • Climate statements and studies • Assessments from the Intergovernmental Panel on Climate Change • Monthly and annual summaries For more information, refer to the WMO Guidelines on Information Management (to be published in 2023 as part of an updated Handbook on WIS)	Optional
4.6.2.2 Documentation	This component refers to a range of textual data that describe various climate-related phenomena or serve as documentation for CDMSs. Some examples may be: **Process and procedural documentation**:	Recommended

	• CDMS technical and user documentation • User handbooks and procedural manuals • Training documentation • NMHS policies and practices applying to data management, such as cloud policies. **Others**: • Text on a web page • Diagrams representing climate processes, such as the Community Earth System Model from the National Center for Atmospheric Research, United States, shown in section 2.1 of this publication • Various climate forecasts and events • Climate processes such as El Niño-Southern Oscillation events. • Other NMHS policies and practices • Various documents and reports • Presentations	
4.6.2.3 Various media	This component covers a range of media used to support various climate-related services on an NMHS website. Some examples may be: • Scanned hard copy climate records • Image portrayal of various climate data, such as an extract from a radar image stored in portable network graphics (PNG) format • Podcasts and video clips used to communicate various climate-related messages • Photographs of various climate-related phenomena	Recommended
4.6.3 Climate software		

<table>
<tr><td colspan="3">
As discussed in the article by the UK Parliament Science and Technology Committee (2011), one of the recommendations of the UK parliamentary review of the Climategate issue was to ensure that climate scientists make available the full methodological workings (including computer codes) used to support their work. An extract is reproduced below.
“It is not standard practice in climate science to publish the raw data and the computer code in academic papers. However, climate science is a matter of great importance and the quality of the science should be irreproachable. We therefore consider that climate scientists should take steps to make available all the data that support their work (including raw data) and full methodological workings (including the computer codes).
This has implications for the effective management of climate data and software in that software source code will also require careful management.“
In addition, it will be necessary to keep track of the time period during which each software version is in operation, as this may also have implications for climate data and climate analysis, and is related to the Provenance metadata requirements for climate data, as outlined in Section 4.3.3.
On a conceptual level, this is similar in a way to the need to have effective observations metadata that describe the maintenance of sensors and stations.
Apart from transparency, proper management of climate software will also assist with operational efficiency.
</td></tr>
<tr>
<td>4.6.3.1 Source code management</td>
<td>
This component deals with managing the source code of the software used to process climate data.
This component should have the following capabilities at a minimum:

Maintain a library of a variety of software source code.
Manage different versions (or branches) of the software concurrently, with the ability to maintain each version independently and to easily backport newer functionalities to an older version.
Easily detect the differences between software versions.

</td>
<td>Rquiered</td>
</tr>
</table>

4.6.3.2 Package management	This component refers to the functionality that facilitates the packaging of software and its configuration for installation on a computer. In addition, the component should facilitate dependency management to ensure that all required supporting software is also installed and configured appropriately at installation time.	Required
4.6.3.3 Environment configuration	This component concerns the functionality that facilitates the recording and management of information relating to any changes to operational software. This includes: • What software was deployed on what server? • What version of the software was deployed? • Details of any configuration changes. • Details of any change made to the operational software. • Details of the decommission of the software at the end of its period of operation, including decommission date.	Recommended
4.6.3.4 Software testing	This component covers the testing of software that is to be deployed to manipulate climate data. This includes: • Details of test plans and individual test cases, including user-acceptance testing. • Details of the test data, database, etc. • Details of test systems and environment. • Details of test results and artefacts, particularly proof that the test data	Recommended

	were not affected by the software or a change to the software.	

4.7 References

4.7.1 WMO

Title	N°	Publication
1961-90 Global Climate Normals CLINO	WMO-No. 847	1998
Challenges in the transition to Automatic Meteorological Observing Networks for long-term climate records	WMO-No. 1202	2017
Commission for Weather, Climate, Water and Related Environmental Services and Applications	WMO-No. 1307	2022
DAYCLI message user guide	NA	2024
Guidance booklet on WMO Maturity Matrix for Climate Data	WMO-No. 1328	2023
Guide to Climatological Practices (4th Edition)	WMO-No. 100	2023
Guide to Hydrological Practices Vol 1: From Measurement to Hydrological information	WMO-No. 168	2008
Guide to Marine Meteorological Services	WMO-No. 471	2018
Guide to the WMO Integrated Global Observing System	WMO-No. 1165	2021
Guidelines for managing changes in climate observation programmes	WCDMP-62	2007
Guidelines for the submission of World Weather Records	NA	2021
Guidelines on the definition and characterization of extreme weather and climate events	WMO-No. 1310	2023
Handbook of drought indicators and indices	WMO-No. 1173	2016
Handbook on CLIMAT and CLIMAT TEMP reporting	WMO-TD 1188	2009
Manual on Marine Meteorological Services	WMO-No. 558	2018
Manual on the High-Quality Data Management Framework for Climate	WMO-No. 1238	2019
Practical help for compiling CLIMAT reports	WMO-TD 1477	2009
Technical Regulations Basic Documents No. 2 Vol I, II and III	WMO-No. 49	2019
The 2022 GCOS ECVs Requirements	GCOS-No. 245	2022
The Role of Climatological Normals in a Changing Climate	WCDMP-61	2007
WIGOS Metadata Standard	WMO-No. 1192	2019
WMO Guidelines on Generating a Defined Set of National Climate Monitoring Products	WMO-No. 1204	2017
WMO Guidelines on Information Management	NA	2023
WMO Guidelines on the calculation of climate normals	WMO-No. 1203	2019
Developing the Climate Science Information for Climate Action		2022
Guide to Instruments and Methods of Observation	WMO-No. 8	2021

4.7.2 Link

Publisher	Title	URL	Current at
Climate Dynamics	Estimating daily climatological normals in a changing climate	https://link.springer.com/article/10.1007/s00382-018-4584-6	Dec, 2023
ClimDex	Climdex	https://www.climdex.org/	Dec, 2023
Climpact	Climpact	https://github.com/ARCCSS-extremes/climpact	Dec, 2023
GCOS	Monitoring Principles	https://gcos.wmo.int/en/essential-climate-variables/about/gcos-monitoring-principles	Dec, 2023
GRDC	Global Runoff Data Centre	https://www.bafg.de/GRDC/EN/01_GRDC/grdc_node.html;jsessionid=8B54714EBEC926A236F366E90A095907.live11294	Dec, 2023
ISO	Geographic Information - Observations, measurements and samples	https://www.iso.org/standard/82463.html	Dec, 2023
ISO	ISO 19156:2011 Geographic Information – Observations and Measurements	https://www.iso.org/standard/32574.html	Dec, 2023
ISO	ISO Glossary	https://www.iso.org/glossary.html	Dec, 2023
NOAA	National Centres for Environmental Information US Climate Reference Network – Instruments	https://www.ncei.noaa.gov/access/crn/instruments.html	Dec, 2023
UCAR	Reanalysis	https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables	Dec, 2023
UN	Terminology Database (UNTERM)	https://unterm.un.org/unterm2/en/	Dec, 2023
W3C	Glossary and Dictionary	https://www.w3.org/2003/glossary/	Dec, 2023
W3C	Overview of the PROV family of documents	https://www.w3.org/TR/prov-overview/#:~:text=The%20PROV%20Family%20of%20Documents%20defines%20a%20model%2C%20corresponding%20serializations,environments%20such%20as%20the%20Web.	Dec, 2023
W3C	PROV-DM: The PROV Data model	https://www.w3.org/TR/prov-dm/	Dec, 2023
Wikipedia	Big Data	https://en.wikipedia.org/wiki/Big_data	Dec, 2023
Wikipedia	Epidemiology	https://en.wikipedia.org/wiki/Epidemiology	Dec, 2023
Wikipedia	Semantic technology	https://en.wikipedia.org/wiki/Semantic_technology	Dec, 2023
WMO	Climate Data Model	https://github.com/wmo-im/tt-cdm/wiki	Dec, 2023
WMO	Observing System Capability and Reference tool	https://oscar.wmo.int/surface/#/	Dec, 2023
WMO	OSCAR – Observing Requirements	https://gcos.wmo.int/en/essential-climate-variables/about/gcos-monitoring-principles	Dec, 2023
WMO	OSCAR – Variables	https://space.oscar.wmo.int/variables	Dec, 2023
WMO	WMO Codes Registry	http://codes.wmo.int/	Dec, 2023
WWW	Web Ontology Language	https://www.ncei.noaa.gov/access/crn/instruments.html	Dec, 2023

5 Climate data management

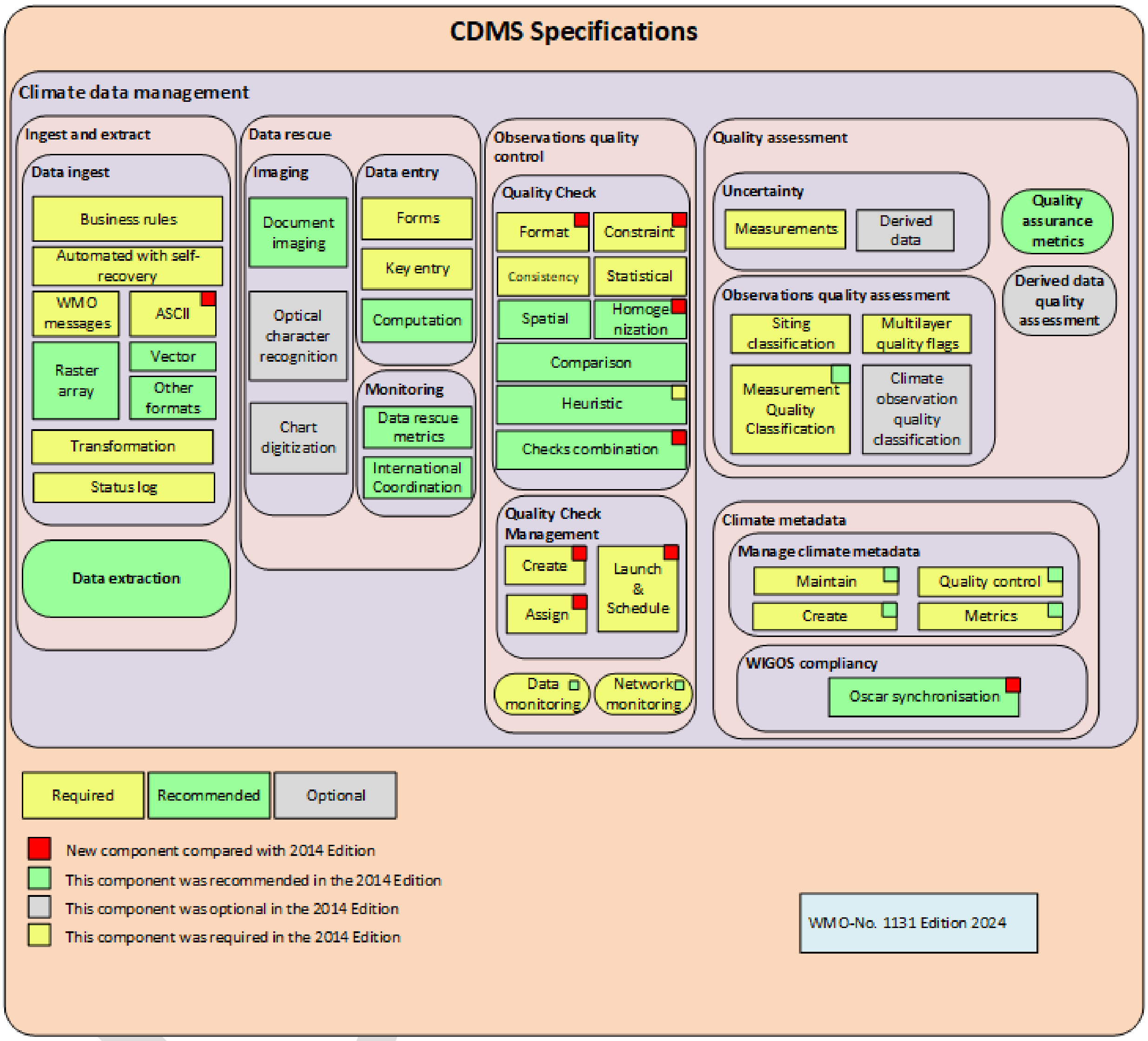


Figure 5.1. Climate data management

The Climate data management component represents the range of functionalities typically required to effectively manage climate data.

5.1 Ingest and extract

This section covers a very broad set of functionalities relating to the capture and initial processing of observation and related data, along with other data types as spatial, gridded, etc.

In essence, this section involves:

- Loading data into or extracting data from the climate database.
- Transforming data as required from one format into another more suitable for data management, analysis and storage.

5.1.1 Data ingest

The capacity of a data management system to be linked to the different data measurements and observations available is essential to allow applications and services. Those links should be as automatic as possible and include monitoring and quality control functionalities.

As illustrated in the following graphic the range of data types to be ingested is very wide and it is often necessary to add dedicated systems to allow data acquisition: system(s) to collect the data, system(s) to process the data before the ingestion into a database.

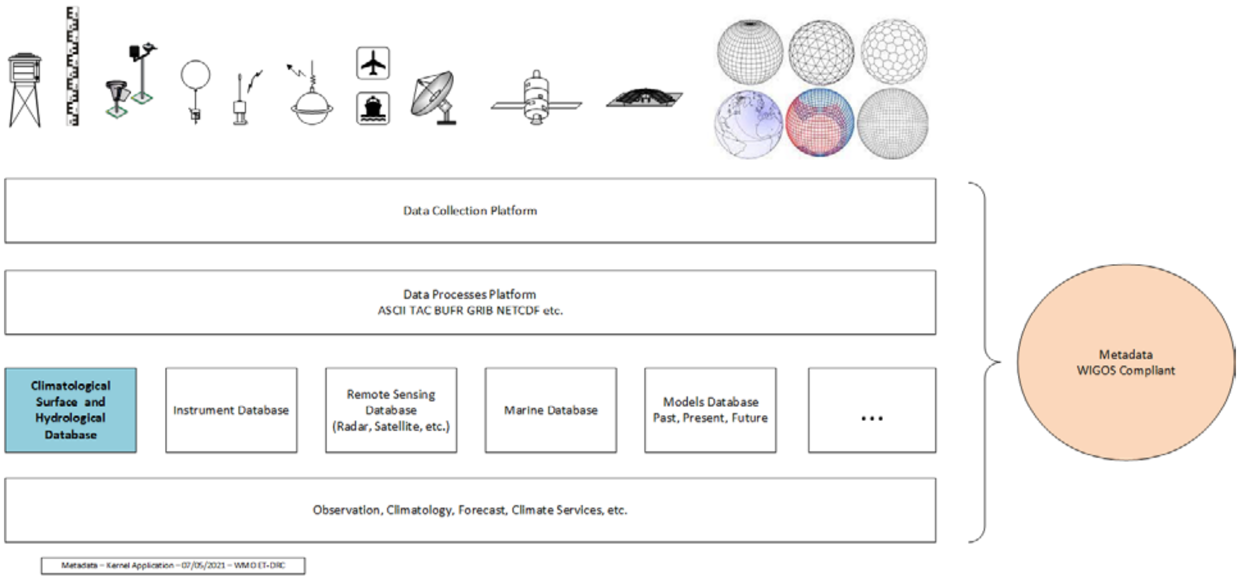


Figure 5.2 Metadata – Kernel Data Flow (Expert-Team on Data Requirements for Climate Services) WMO-No. 100, 2023 Edition

Note that data ingest may take the form of electronic transfers from measurement systems such as Automatic Weather Stations, from upstream databases within the data flow system, from the GTS / WIS, and from key entry from paper forms. The requirements below are applicable to all these forms.

5.1.1.1 Business rules	This component supports a wide range of user defined business rules that govern how data are ingested into the climate database. Some examples (for observations data) are:	Required

	• Action required when new measurements are to be ingested, but a record already exists in the database for that time period. • Should the new record replace the current record in the database or should the new record be rejected? • There is potential for data that have not been quality controlled to overwrite perfectly good quality-controlled data. An example is a message that is reingested and the ingest process does not take into account the possibility that the data already exist in the database and that they have been modified. • Action required when a message arrives for ingest but the message type is not appropriate according to the observations metadata on record for that station. • Action required should a message arrive containing an observed value that is outside the accepted bounds for a given phenomenon. For example, a message contains a value of 90°C for temperature, where the maximum acceptable temperature is 60°C. The same applies to “impossible” values such as negative rain. • Action required should a message arrive that is of a lower order of precedence to one that has already been ingested for the same time period and station. For example, the priority level given to records being ingested may relate to the method of data acquisition. A record that has been keyed in via a quality assurance process may be given a higher priority than a record acquired via a real-time message ingest.	

	Business rules for data ingestion are crucial for a CDMS and it is important that such rules be set up according the NMHSs practices and needs, and in line with WMO requirements.	
5.1.1.2 WMO messages	This component allows for the import of data from a range of WMO message formats, including Traditional Alphanumeric Codes (TAC) and Table Driven Code Forms (TDCF). As both historical and current data will need to be imported, this component should be able to work with data in a wide variety of past, present (and future) data formats. Some examples are: • Binary: ✓ BUFR (FM 94) ✓ GRIB • Alphanumeric: ✓ CREX (FM 95) ✓ SYNOP (FM 12) ✓ TEMP (FM 35) ✓ SHIP (FM 13) ✓ METAR (FM 15) ✓ CLIMAT (FM 71) ✓ Etc. Note: While TAC formats are being phased out, support for them will still be required by this component to support the ingest of historical data. The same applies to obsolete message types such as CLIMAT TEMP (FM 75). For more information, see: Manual on Codes WMO-No. 306	Required
5.1.1.3 Vector	This component supports the import of a series of vector spatial formats.	Recommended

	For example: • Shapefile • Geography Markup Language (GML) (see [OGC GML web page]) • Keyhole Markup Language (KML) • GeoJSON, JSON, OGC geoPackage, GeoDatabase, XML, etc.	
5.1.1.4 **Raster array**	This component supports the import of a series of raster array spatial formats. For example: • Network Common Data Form (CF-netCDF) • Hierarchical data format (HDF) • ArcInfo ASCII • GeoTIFF • Etc.	Recommended
5.1.1.5 **ASCII**	This component addresses the importation of a variety of ASCII formats. Specific domains have their own formats and it is critical that the system have the capacity to ingest global, regional and national data from these domains in ASCII formats. Those ASCII formats are very different from one to another, so the system should be able to easily create specific ingestion processes (past, current and future ASCII formats). For example, be able to ingest: • Data from the National Centers for Environmental Information (NCEI) like the Global Surface Summary of the Day (GSOD), the Global Historical Climatology Network (GHCN), the Global Hourly Integrated Surface Database (ISD), the International Surface Pressure Databank;	Required

	• Globally the International Maritime Meteorological Tape (IMMT) and the International Maritime Meteorological Format (IMMA); • Regionally, the Word Weather Records; • Nationally, the formats of historical data used in the country, or the formats of other CDMSs such as CLICOM, CLIMSOFT, MCH, CLIDE, etc.	
5.1.1.6 **Other formats**	This component covers the import of a range of other formats. For example: • Photographs (PNG, JPEG, TIFF, etc.) • Scanned documents • PDF files • Data managed in spreadsheets • Tabular formats, such as the import of data from a relational database management system • New data type of JSON Note: Scanned documents, pdf files may be stored in subsidiary storages, but the CDMS should link to or refer to these.	Recommended
5.1.1.7 **Status log**	This component concerns the recording of each ingest activity status in order to: • Monitor the ingest job status, and allow performance statistics of the ingest process. • Automatically recover failed ingests. • Record warning and other error messages to enable intervention if required, for example if expected data are not received.	Required

5.1.1.8 Automated with self-recovery	This component supports the automated ingest of a range of ingest types (particularly WMO messages and data from automatic weather stations). The component also allows for the automatic recovery of ingest tasks in the event that a task fails either entirely or part way through an ingest. This could be due to a number of reasons, including: • Corrupted messages • Network failures • Hard disk failures • Database failures • Upstream data flow disruptions.	Required
5.1.1.9 Transformation	This component supports the transformation of an ingest record. This may include: • Transforming data from one format to another (e.g. unit conversion according the governance in place for data rounding as mentionned in 3.1.6.3). • Transforming codes into formats more suitable for the destination climate database. For example translating SYNOP codes values into actual measurements such as visibility in km. • Correcting records that have been abbreviated in accordance with accepted local observation practice. Such data processes are candidates to be captured by lineage metadata. See the component 4.3.3 on Provenance.	Required

5.1.2 Data extraction

5.1.2.1 **Data extraction**	This component allows data to be extracted from the climate database in accordance with NMHS data policy and governance processes. Data may be transformed into a wide range of formats as described in the subsection on data ingest (5.1.1). Note: This component is only intended for advanced users who have an intimate knowledge of the climate database, its data structures, the relevant data policies and the appropriate use of quality flags and other aspects in order to perform one-off data extraction activities. End-user data extraction is intended to be constrained to defined data types via the climate data delivery services components (Chapter 8), using components under Chapter 7, such as: Tables and charts, Integrated search of climate data and Data download. End users should not be given direct access to the database itself, because of the risk of data corruption (accidental or intentional).	Recommended

5.2 Data rescue

5.2.1 Imaging- & digitization

5.2.1.1 **Document imaging**	This component supports the functionality required to digitally capture a physical document and store the resultant file and associated discovery metadata, either within the climate database, or via an integrated subsidiary database.	Recommended

	Some examples of the types of documents to be digitally captured are: • paper observation forms • microfiche/microfilm • Relevant observations metadata documents such as instrument calibration reports • Technical manuals • Site location plans and sections For more information, see: • Guidelines on Best Practices for Climate Data Rescue (WMO-No. 1182), • Guidelines for Hydrological Data Rescue (WMO-No. 1146), • Best Practice Guidelines for Climate Data Rescue from the Copernicus Climate Change Service	
5.2.1.2 **Optical character recognition**	This component provides the functionality required to digitally capture data stored in scanned documents such as hand written and/or typed meteorological observation forms. While not very effective at present on hand-written forms, future advances in machine-learning/artificial intelligence may change this situation	Optional
5.2.1.3 **Chart digitization**	This component refers to the capacity to digitize data from recording cards such as those used with a Campbell-Stokes sunshine recorder, thermograph, barograph, pluviograph, or other meteorological instrument. The typical functionality required for this component would be to: • Scan a physical recording chart (or card) using the Document imaging component (5.2.1.1). • Analyse the image of the chart.	Optional

	• Extract numeric points from the chart, and provide a numerical value for those points. • Store the resultant data as a time-series in the climate database.	

5.2.2 Monitoring

5.2.2.1 Data rescue metrics	This component maintains metrics relating to the capture of historical observations data via a data rescue project. These may contain: • Name and brief description of data rescue project • Countries where activity is taking place • Point of contact information for project • Types of data rescued • Summary and per cent digitized • Summary and per cent scanned • Summary and per cent scanned but not digitized • Summary and per cent undigitized	Recommended
5.2.2.2 International Coordination	This component describes the need to exchange information between applications regarding Data Rescue, including the kind of information described in Section 5.2.2.1. It should be used for example to synchronize the information contained within International Data Rescue portals, such as the ones from WMO I-Dare, and the Copernicus Climate Change Service, with the CDMS. See Data Rescue Portals:	Recommended

	• https://idare-portal.org/ • https://datarescue.climate.copernicus.eu/map The need for this standard has been highlighted in the report "Data Rescue Activities 2021" available at : https://insitu.copernicus.eu/library/reports/copy_of_20211217_EEA_Report_DataRescue_v8_sent.pdf It should also be noted that such data portals should also contain information on data to be rescued and where to find it, and on available software and initiatives to support data rescue, including initiativeatives such as Oldweather, IEDRO (for International Environmental Data Rescue Organisation), etc., and software such as the Weather Wizards software developed by IEDRO for digitizing strip charts (pluviograph, barometric traces, etc.).	

5.2.3 Data entry, also known as key entry

This subsection covers the functionality required to enable an appropriately trained and authorized person to manually enter data into the climate database.

Typically, this functionality is tightly controlled according to NMHS data governance processes.

Some issues to consider are:

- Data entry staff should only be able to add data to or edit data in the climate database under programme control, with appropriate safeguards in place to protect the integrity of the climate database.

<table>
<tr><td colspan="3">

- Any functionality that provides write access to the database should also include an audit function to allow an independent review of database changes.
- One example could be the use of database triggers that write the details of a transaction, including the previous values, into a separate set of audit tables.
- Another approach could be to ensure that the data entry process creates an interim data file that is then entered into the database via data ingest processes, bypassing the need for direct access to the database.
- NMHS data policy may enforce the need for double (or multiple) entry practices, where two or more operators key in data for the same form, independent of each other, to detect and minimize key-in errors. This practice involves having provision to store the multiple values at least in the short term.
- Careful consideration should be made to ensure that an organization has very effective IT security and monitoring in place prior to allowing key-in access via the Internet, as might be the case for citizen-science digitization projects, for instance. Most organizations will not have suitable controls in place.
- NMHS data policy should provide guidelines as to appropriate data quality considerations applied to data that are manually entered.

</td></tr>
<tr><td>5.2.3.1 Forms</td><td>

This component covers:

- The visual design of a form.
- The software logic that controls the data key-in process.
- The mapping of fields in the form with appropriate records and tables within the climate database.
- Ensuring that the integrity of the climate database is protected by validating data before they are added to the database.

The component should also support:

- A custom definition of user input forms that mimic traditional meteorological forms (including the language where appropriate). Where data have been recorded in a large number of different versions of a form, it may be more efficient to design an operator-customisable form by, e.g., drag and drop functionality, and store the templates so produced as an operational suite.

</td><td>Required</td></tr>
</table>

	• Efficient and effective data entry that minimizes operator fatigue and automatically calculates appropriate values. The component should provide adequate support for monitoring the validity of data that are entered. Some examples are: • Performing data quality consistency checks of the data to be entered. These checks and the appropriate values are to be customizable according to NMHS data policy and governance processes. • Ensuring that appropriate data types and context are entered for each field. • The component should alert the operator to any doubtful entries detected, providing appropriate advice as per NMHS data policy guidelines.	
5.2.3.2 **Key entry**	This component provides the functionality to support manual key-in of meteorological data.	Required
5.2.3.3 **Computation**	This component allows for the automatic derivation of parameters at key-in. Such computation should be customizable according to NMHS data policy and governance processes. Some possible scenarios where this functionality may be used are: • The computation of a value for relative humidity after the values for dry-bulb temperature and dewpoint have been entered. • Decoding shorthand codes and replacing them with appropriate values	Recommended

	• Modifying units, e.g. from historical units (Torr or mmHg into Hpa). As from 4.3.3 the system should ensure Provenance and Lineage of such data modification.(4.3.3)	

1

5.3 Observations quality control

The purpose of this component is to ensure that all data have been checked for accuracy, and that a record is maintained of what checking was performed, the outcome of the checks and when the checking (and possible changes to the data) took place. The checked data value should be assigned a quality flag or by other means indicate the confidence placed in it.

In line with Provenance requirements (4.3.3) traceability of the data and its QC indicator should be retained.

5.3.1 Quality checking

This component involves checking data via series of checks to ensure that it is a reliable representation of reality, and that it has not been corrupted by any process. For practical purposes, the CDMS should have the ability to scan the data automatically and identify any potentially suspect observations. It is common (and recommended) practice to then assess suspect values via application of a set of tools, or an expert system.

For Marine data, Quality checking is carried out via various processes within the nodes of the Marine Climate Data System (MCDS), as described in the Manual on Marine Meteorological Services (WMO-No. 558), and the Guide to Marine Meteorological Services (WMO-No. 471). More work is required on gathering and harmomizing Marine QC and Surface QC at the NMHS level.

For more information, see:

- *Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications* (WMO-No. 1269, 2021)
- NORDKLIM, *Quality Control of Meteorological Observations Automatic Methods Used in the Nordic Countries* (https://www.yumpu.com/en/document/view/40570385/quality-control-of-meteorological-observations-smhi)
- *Guide to Climatological Practices* (WMO-No. 100)

<table>
<tr><td colspan="3">

Guide to the Global Observing System (WMO-No. 488), Appendix VI.1 Data quality control, and Appendix VI.2 Guidelines for quality control procedures applying to data from automatic weather stations
Guide to the Integrated Processing and Prediction System (WIPPS)WMO (WMO-No. 305), Chapter on Quality control procedures

</td></tr>
<tr><td>5.3.1.1 Format checks</td><td> This component checks if the data format conforms to the expected format. For example, it could detect that an amount of precipitation with a precision of thousandth of millimeter, instead of the traditional tenth of millimeter, would be suspected as an error in the format. </td><td>Required</td></tr>
<tr><td>5.3.1.2 Constraint checks</td><td> This component verifies if the data is physically in the right limits (having regard to climatic conditions at the local, national, regional, global scales). For example, at a global level, a temperature falling outside the range -110 °C and +80 °C is almost certainly wrong. </td><td>Required</td></tr>
<tr><td>5.3.1.3 Consistency checks</td><td> This component covers a range of tests to ensure that inconsistent, unlikely or impossible records are either rejected or flagged as suspect. A manual investigation may then assess the validity of the suspect values. This component includes the concepts of internal, temporal and summarization consistency checks. WMO-No 1269 contains some examples, but many more are possible. Some examples are: Is the minimum temperature lower than the maximum temperature?
Is the maximum temperature within the historical range for maximum temperatures for a given station?</td><td>Required</td></tr>
</table>

5.3.1.4 **Statistical checks**	This component covers a number of tests that statistically analyses historical data to detect inconsistent, unlikely or impossible values and flag them as suspect. A manual investigation may then assess the validity of the suspect values. Some examples are: • Climate tests that highlight extreme climatic values, such as a record maximum air temperature. • Flatline tests, where a constant value exceeds a specified limit in a time series, for example when the station air temperature remains constant for 12 hours. • Spike tests, conducted in a time series to identify data spikes exceeding a specified limit, for example when a three-hourly air temperature observation is at least 10 °C colder than all others during the day. • Rapid change tests conducted in a time series to identify improbably rapid changes exceeding a specified limit. For example, when a 100 cm soil temperature suddenly changes in consecutive 3-hourly observations from a relatively stable 22°C to 38°C and remains there.	Required
5.3.1.5 **Spatial checks**	This component covers a range of spatial tests to detect spatially inconsistent, unlikely or impossible records and flag them as suspect. A manual investigation may then assess the validity of the suspect values. Some examples are: • Comparing the results of a time series of observations at a given station with those at nearby stations. • Using a Barnes interpolation or similar mapping analysis to derive spatial patterns against which anomalous	Recommended

	and possibly erroneous station values stand out.	
5.3.1.6 **Homogenization checks**	This component utilizes homogenization techniques for the detection of suspicious data. See also Guidelines on Homogenisation (WMO-No. 1245)	Optional
5.3.1.7 **Comparison checks**	This component covers a series of tests that use and cross-reference data from a number of sources to detect suspicious observations. Some examples of data sources that may be cross-referenced are: • Observations data or maps showing daily precipitation at a station • Radar data covering the station • Synoptic forecast charts • Satellite imagery (for instance when significant precipitation is recorded, but satellite imagery suggests no cloud) i.e. Aerological diagram traces. • Numerical Weather Predictions • Reanalysis products	Recommended
5.3.1.8 **Heuristic checks**	This component refers to a set of tests that rely on experience and knowledge of observation processes, techniques and instrumentation to detect inconsistent, unlikely or impossible records and flag them as suspect. A manual investigation may then assess the validity of the suspect values. Artificial intelligence could be used in the future for error detection. Some examples are problems typically caused by: • Inexperienced operators • Instruments that are not or are incorrectly calibrated, or when the measurement exceeds the instrument measuring capacity. In certain	Recommended

	circumstances such as failure of an instrument, overflowing raingauge, blockage (e.g. due to snow) the value/total recorded represents a known minimum. For instance, when a manual raingauge overflows, this means that at least 200mm has fallen (assuming the observer has not neglected to tip out the raingauge). Such lower-bound information may still provide very valuable data for a range of climate applications, including hydrological • Operator behaviour or organizational policy, for example not recording rainfall data over a weekend period and aggregating the results on the following Monday, or otherwise reporting data on the wrong day • Known deficiencies in observers handling data such as evaporation-related observations • Changes over time caused by changes at an observation site. For example, a shift in the magnitude of wind recorded from a specific direction may be an indicator of a problem at the site location, such as a new building structure or trees obstructing the flow of the wind in that direction.	
5.3.1.9 **Checks combination**	This component refers to the capacity of the user to create checks that are the result of a combination of the above checks. e.g. values considered as suspect after a Spatial check are automatically checked with Statistical checks and then by Comparaison checks against satellite and radar data. Personnel will typically review and consider a wide range of data in their investigations, such as raw records, synoptic charts, satellite imagery, radar, metadata e.g.	Recommended

	siting classification, and other sources as in component 5.3.1.2 above.	

5.3.2 Quality Check Management

This component describes the overall management of the quality checking process. It includes the capacity to create and modify data quality checks, to assign quality control parameters to certain checks and to launch the process of data quality checks with scheduling properties.

5.3.2.1 Create and Modify	This component defines what quality control checks should be applied on the data, and how they should be applied. It should be able to build format, constraint, consistency, statistical, spatial, homogenization, comparison and heuristic quality checks. It should allow scope for new tests to be introduced or the modification of existing ones (for instance, it may be necessary to modify certain tests if monitoring shows too many false positives are created). Such modifications should be subject to rigorous version control and documentation.	Required
5.3.2.2 Assign variables	This component assigns specific set of quality controls to specific variables, specific stations and specific time periods. Example: to assign to the Temperature at 2 meters a set of specific quality controls (consistency, statistical, spatial, etc.), that take into account specific location, altitude and season. E.g. that the temperature of all stations that do not exceed the altitude of 100 meters in a specific area of the country must be between 8°C and 36°C during summer time.	Required

5.3.2.3 Launching and Scheduling	This component is responsible for the launching of the data quality control process. It should launch such processes as scheduled by NMHSs for example: • Each hour, to check the data newly ingested into the system against a list of data quality checks • Each day to launch spatial quality control on daily extreme and accumulated data after the global daily computation process of the system • Provide quality Control on historic data, e.g, on observations/measurements between 1932 and 1956 newly ingested into the system as a result of Data Rescue processes.	Required

Figure 5.3. Example of representation of data acquisition albeit on a global scale (WIGOS data monitoring system) illustrating a form of network monitoring

5.3.2.4 **Network monitoring**	This component keeps track of data received from an observational network and the frequency and type of errors at constituent stations in order to monitor network performance and potentially detect and alert data and observation managers of possible problems. Some examples are: • Statistics showing the current state of data ingestedz compared with historical averages. • Records not received for a station or sensor could indicate a potential issue with the instrumentation, IT equipment, software or IT network. • Specific fields or records not loaded into the database could be an indication of a systemic error such as an undetected software error. • Records received in advance of the observation time may highlight training issues or time drift for a Automatic Weather Station. • Loss of data due to corrupted synoptic message types. For more information, see: • *Guide to the Global Observing System* (WMO-No. 488), section 3.1.3.14 Network performance monitoring • *WMO-No. 1269*, Chapter 2.3	Required

5.3.3 Data monitoring

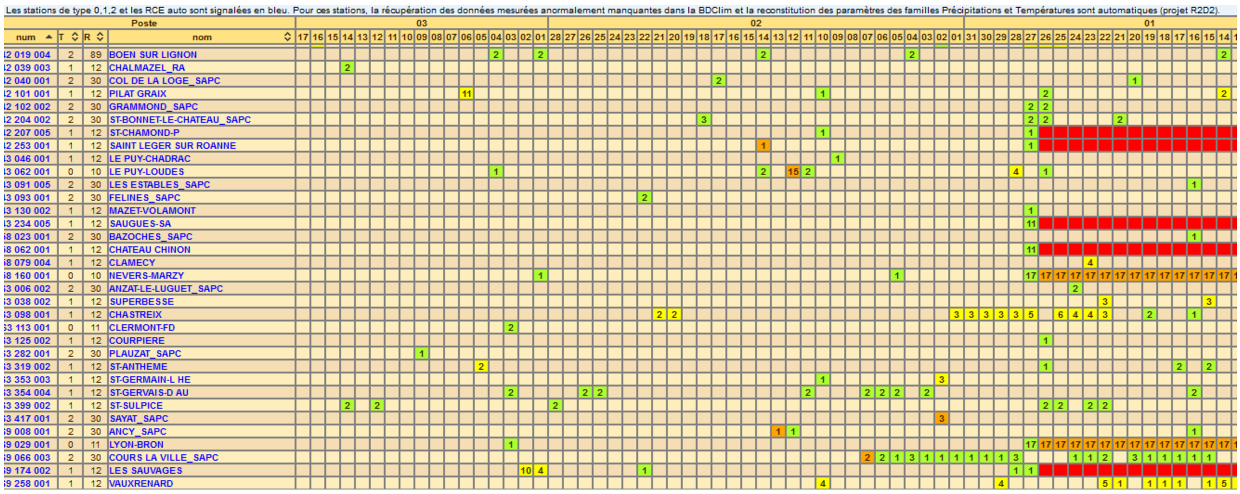


Figure 5.4. Example of representation of missing data over 70 days on a specific region, updated every 15 minutes

5.3.3.1 Data monitoring	This component monitors ingested and derived observation data to detect and resolve potential systemic issues. Verifying, in real-time and in past-mode, the **completeness** of the database is one of the major tasks of this component. The component should check if all data expected to be ingested is physically present in the database. Another major tasks is to look for systemic patterns of error at specific stations, which may indicate inadequate maintenance or observer training issues. Another necessary function is the verification of the database coherency after any data modification. For example, NMHSs typically have quite long data quality control processes, where data may be modified over a period lasting several months. In the intervening period, normal business processes create considerable amounts of derived data, such as:	Required

	• Daily, weekly, monthly, seasonal and annual summaries for a range of observation variables. • Statistical gridded datasets that represent the distribution of observation variables over large areas for similar time periods. As soon as observation data are modified through normal quality control processes, derived data may become invalid. This is why data monitoring processes are required to monitor data changes and, where necessary, reconstruct affected derived datasets. A Notification service component may be beneficial to those who are using a particular dataset. NMHSs should encourage users of a datasets to register themselves to the service to receive updates about availability, errors and update to the data used in their workflows. Another example is the generation of a number of metrics to assist with climate data management activities.	

1

5.4 Quality assessment

5.4.1 Observations quality assessment

This component refers to the processes implemented to help NMHSs assess the quality of observations used by their organization. It covers all stages, from the observation site and expertise level of personnel to the final product distributed to users.

The aim of this subsection is to move towards a more objective way of defining the quality of observations data.

5.4.1.1 Siting classification	This component refers to the processes, software, governance mechanisms and analysis that classify sensors according to the rating scale described in the *Guide to Instruments and Methods of Observation* (WMO-No. 8), Annex 1.D Siting classifications for surface observing stations on land.	Required
5.4.1.2 Measurement Quality Classification	This component refers to the processes, software, governance mechanisms and analysis that classify sensors according to their quality, maintenance and calibration status. *Guide to Instruments and Methods of Observation* (WMO-No. 8), Annex 1.G. Measurement quality classifications for surface observing stations on land.	Required
5.4.1.3 Multilayer quality flags	This component refers to the processes, software, governance mechanisms and data analysis used to understand and enumerate the quality flags of specific aspects of a specific record of data. This will facilitate:	Required

	• Future applications that requires data of a specific quality flag value. Data for climate change analysis will normally require a higher level of confidence, and in general transparency, than that for other climate applications. • Communication on the assessed quality of records. The best description to date on how to define this classification may be found in the *Guide to Climatological Practices* (WMO-No. 100), section 2.4.5. This reference describes a way of flagging quality based on a combination of: • Data type (original, corrected, reconstructed or calculated) • Validation stage • Acquisition method Note that this is separate from, but complementary with, quality flags assigned following a quality control process, such as the quality control flags set out in the DAYCLI message (Code Table 031 021 Associated field significance, FM 94 BUFR). While the classifications are relevant and relate to the perceived quality of a record, they do not allow for an explicit comparison of data of similar perceived quality. For example, the subsection on quality management (5.3.1) describes a series of classifications of tests (the details of which are provided in the quoted reference). Objective quality classifications are required to support a consistent approach within the global WMO community so that data can be: • Objectively compared to ensure that data of similar quality can be compared and analysed as required. • Stored and easily retrieved from a climate database. For Provenance	

	reasons, Organizations need to retain observations at multiple levels of quality, from the raw observation through various quality checking and analysis processes in order to demonstrate the true lineage of a record and explain and justify the changes made to the raw observations. Note: A more objective approach to determining this classification for the global WMO community is planned, with a Guidance document to be developed.	
5.4.1.4 **Climate observation quality classification**	This component refers to the processes, software, governance mechanisms and data analysis used to understand and enumerate the quality of a specific record of data relative to an objective index. This index will need to combine a number of criteria relevant to data reliability and quality. Note: This index has yet to be created. For the purposes of this publication, it is called the climate observation quality classification. However, this name may change. It is envisioned that this index will need to take into account a number of factors, including: • Siting classification • Measurement Quality Classification • Sensor reliability • Calibration • Maintenance regime • Uncertainty inherent in observations • Observation quality control processes • Multilayer quality flags • Lineage • Homogeneity • Stewardship Maturity Assessment	Optional

	• Other appropriate factors	

5.4.2 Derived-data quality assessment

5.4.2.1 Derived-data quality assessment	This component refers to the processes, software, governance and data analysis processes used to understand and enumerate the quality of derived data relative to an objective index.
	
	There are many factors that can influence the quality of derived data. Some issues to consider are:
	• What is the quality of the source data?
	• What algorithms have been applied to the source data to arrive at the derived data?
	• What is the impact of these algorithms on the quality of the derived data?
	• If the derived dataset is spatial, how has the positional location of the data been derived?
	    ✓ What is the quality of the source spatial data?
	    ✓ What is the impact of the algorithms used to spatially distribute the data on the positional accuracy of the derived data?
	
	For more information, see also the Derived data component (5.4.4.2).
	
	Note: This index has yet to be created. For the purposes of this publication, it is called the derived-data quality assessment. However, this name may change.
		Optional

5.4.3 Quality assurance metrics

5.4.3.1 Quality assurance metrics	This component refers to the processes, software, governance mechanisms and analysis used to monitor the performance of quality assurance processes. Such monitoring will allow observational network managers and climate data specialists to validate the performance of quality assurance software and processes, and assist in the enhancement of quality control algorithms. This can be done, for example, by reviewing automatically generated reports that: • Summarize observational errors, and types of errors, detected by each quality assurance test. • Compare the performance of current quality assurance metrics with historical averages. • Summarize data that have undergone quality control with the tests applied for each datum. • Summarize false positives or false negatives to assist with tuning of QC algorithms. • Summarize data modified, data recovered, data never modified (raw data), data awaiting decision, etc..	Recommended

1

2

5.4.4 Uncertainty

This subsection refers to the processes, software, governance processes and data analysis used to understand and record the uncertainty inherent in the data.

As noted in the OGC Abstract Specification: Geographic Information – Observations and measurements, all observations have an element of uncertainty:

"The observation error typically has a systemic component, which is similar for all estimates made using the same procedure, and a random component, associated with the particular application instance of the observation procedure. If potential

<table>
<tr><td colspan="3">
errors in a measurement are important in the context of a data analysis or processing application, then the details of the method of observation which provided the estimate of the value are required."
Some sources of uncertainty are of more importance than others in a reference time series. Random, uncorrelated, uncertainties in a single observation (e.g. as defined by an observation tolerance) are of limited importance in a longer-term perspective. However, correlated errors, including biases that persist over an extended period of time, can be a much greater contributor to uncertainly in the longer term.
This functionality will support:

Future statistical analysis that takes into account the uncertainty inherent in data.
Communication of data uncertainty.
For a description of Uncertainty as it applies to climate observations, see Chapter 4 of paper on Climatological reference stations: definitions and requirements (forthcoming).

For more information, see Wikipedia articles on:

Uncertain data
Uncertainty (broad assessment but concentrate on measurement uncertainty)

</td></tr>
<tr>
<td>5.4.4.1 Measurements</td>
<td>
This component refers to the processes, software, governance mechanisms and data analysis used to understand and record the uncertainty inherent in observation measurements and processes.
For more information, see:

Guide to Instruments and Methods of Observation (WMO-No. 8), 2023 Edition, Annex 1.A Operational measurement uncertainty requirements and instrument performance
The WMO Observation requirements database

The WMO observing requirements database (OSCAR) includes three limits for most meteorological variables: (a) the goal: an ideal requirement; (b) the breakthrough: an intermediate level between threshold and goal which, if achieved, would result in
</td>
<td>Required</td>
</tr>
</table>

	a significant improvement for the target application and (c) the threshold: the minimum requirement to be met to ensure that data are useful..	
5.4.4.2 **Derived data**	This component refers to the processes, software, governance mechanisms and data analysis used to understand and record the uncertainty inherent in data that have been derived from observations data. E.g. many factors can contribute to the uncertainty inherent in gridded derived data. Some examples are: • Uncertainty inherent in the source observation data. • Uncertainty inherent in the location of sensors/stations used to generate the grids. • The relative accuracy of the algorithms used to generate the derived data. • The precision of variable data types used in the software that generates derived data. • The inherent spatial variability of the parameter in question, and its relation to topography. For instance, rainfall in a convective regime, and minimum temperature in a topographically-complex area, will have high uncertainties. It is also worth noting that a number of these factors may propagate through the data derivation process.	Optional

1

5.5 Climate metadata

5.5.1 Manage climate metadata

This subsection refers to the processes, software, governance mechanisms and data analysis required to effectively manage climate metadata, which include metadata on observations, dataset discovery and data provenance.

Note: This subsection is deliberately kept generic in this version of the CDMS Specifications, with all types of climate metadata bundled together. It includes Observation metadata and Discovery metadata which have been both defined, and Provenance metadata for which a WMO standard is still in its early stages noting that a generic PROV standard already exist (see sub-component 4.2.2.4)

For more information on climate metadata, see section 4.3 of this publication.

5.5.1.1 **Create**	This component refers to the processes, software and governance processes needed to effectively and efficiently create climate metadata.	Required
5.5.1.2 **Maintain**	This component covers the processes, software and governance mechanisms required to effectively and efficiently maintain climate metadata.	Required
5.5.1.3 **Quality control**	This component deals with the processes, software and governance processes needed to effectively and efficiently assess and control the quality of climate metadata. More work is required to provide effective guidance on this component.	Required
5.5.1.4 **Metrics**	This component refers to the processes, software and governance processes required to effectively and efficiently maintain metrics relevant to climate metadata. Some examples are: • Which stations or sensors do not have observations metadata records?	Required

	• Which datasets do not have discovery metadata records? • Which instruments meet, respectively, the Goal, Breakthrough and Thresold levels according to the OSCAR requirements? More work is required to provide effective guidance on this component.	

5.5.2 WIGOS compliancy

5.5.2.1 OSCAR Synchronisation	This component refers to the observations metadata that WMO Members are required to make available internationally through the Observing Systems Capability Analysis and Review (OSCAR) tool. It generally gives the capacity to the CDMS: • to synchronise its obervations metadata with OSCAR • to monitor the level of completeness of OSCAR in line with the metadata present in the CDMS • to monitor the level of completeness of the observations metadata contained in the CDMS with the OSCAR requirements.	Recommended

5.6 References

5.6.1 WMO

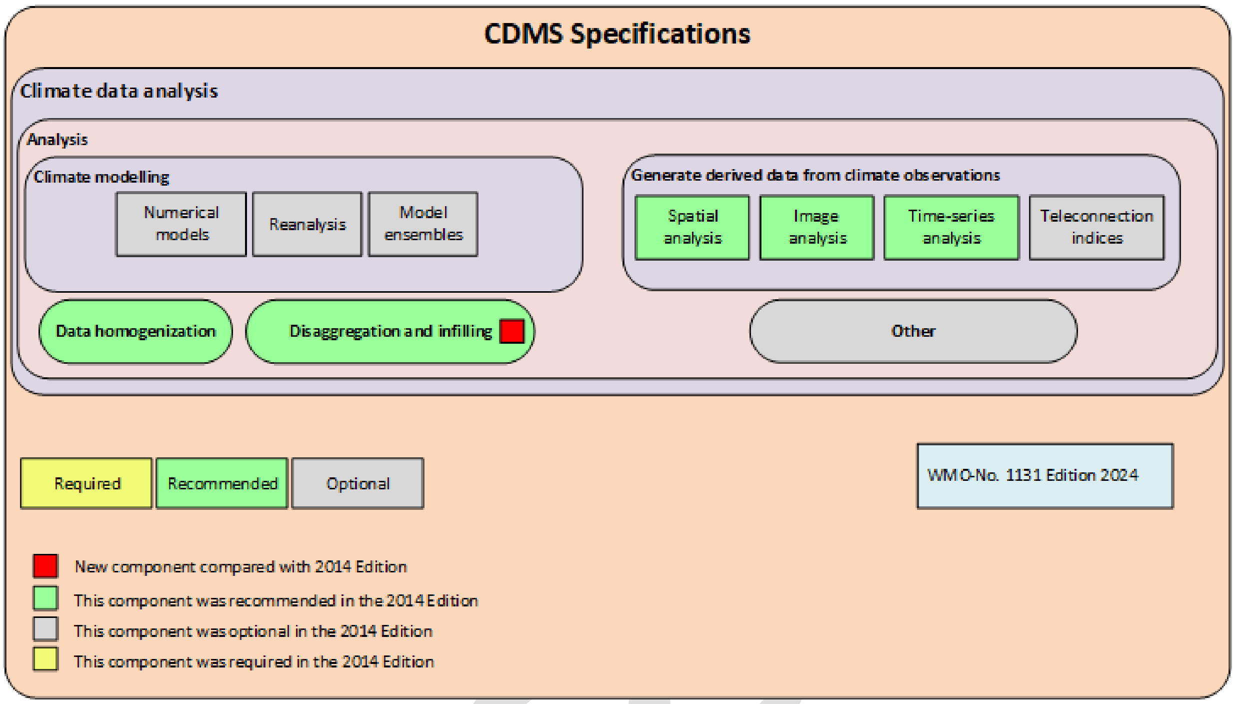
Title	#	Publication
Guide to the Global Observing System	WMO-No. 488	2013
Guidelines on homogenisation	WMO-No. 1245	2020
WMO Guidelines on surface station climate data Quality Control and Quality Assurance for climate applications	WMO-No. 1269	2021
Guide on the Global Data-processing System, Chapter 6 - Quality control procedures	WMO-No. 305	2021

Manual on Codes	WMO-No. 306	2019
Guidelines for hydrological data rescue	WMO-No. 1146	2014
Guidelines for best practice on climate data rescue	WMO-No. 1182	2016

5.6.2 Links

Publisher	Title	URL	Current at
Copernicus C3S	Best Practice Guidelines for Climate Data Rescue from the Copernicus Climate Change Service 2019/C3S_311a_Lot 1_Met Office/SC2	https://climate.copernicus.eu/sites/default/files/2020-02/BestPracticeGuidelines_ClimateDataRescue_0.pdf	Dec. 2023
Copernicus C3S	Copernicus Climate Change Service	https://datarescue.climate.copernicus.eu/map	Dec. 2023
Copernicus C3S	Data Rescue Activities 2021	https://insitu.copernicus.eu/library/reports/copy_of_20211217_EEA_Report_DataRescue_v8_sent.pdf	Dec. 2023
Norwegian Met Inst	NORDKLIM, Quality Control of Meteorological Observations Automatic Methods Used in the Nordic Countries	https://www.yumpu.com/en/document/view/40570385/quality-control-of-meteorological-observations-smhi	Dec. 2023
OGC	OGC Geographic Markup Language Web page	https://www.ogc.org/standard/gml/	Dec. 2023
Wikipedia	Uncertain Data	https://en.wikipedia.org/wiki/Uncertain_data	Dec. 2023
Wikipedia	Uncertainty	https://en.wikipedia.org/wiki/Uncertainty	Dec. 2023
WMO	The International Data Rescue (I-DARE) Portal	https://idare-portal.org/	Dec. 2023

6 Climate Data Analysis



The Climate Data Analysis functional group represents the range of functionality that is typically required to effectively analyse Climate data.

6.1 Analysis

This component describes a range of processes used to analyse climate data.

More work will be required to expand on this component in future revisions of this specification.

6.1.1 Climate Modelling

6.1.1.1 Numerical Models	This component represents the software, processes and governance that provide Earth System Models, e.g. General Circulation Models (also known as Global Climate Models). NOAA defines a GCM as:	Optional

	"A global climate model (GCM) is a complex mathematical representation of the major climate system components (atmosphere, land surface, ocean, and sea ice), and their interactions. Earth's energy balance between the four components is the key to long-term climate prediction.The main climate system components treated in a climate model are: • *The atmospheric component, which simulates clouds and aerosols, and plays a large role in transport of heat and water around the globe.* • *The land surface component, which simulates surface characteristics such as vegetation, snow cover, soil water, rivers, and carbon storing.* • *The ocean component, which simulates current movement and mixing, and biogeochemistry, since the ocean is the dominant reservoir of heat and carbon in the climate system.* • *The sea ice component, which modulates solar radiation absorption and air-sea heat and water exchanges."* Note: Most NMHS would not manage infrastructure and software required to support this component. However, the data outputs from such components may be useful and are available for NMHS use via a range of sources, as refered in the Guide to WIPPS (WMO-No.305) and the WMO WIPPS web portal at: https://wmo.maps.arcgis.com/apps/dashboards/7c3d45e5003a417988bad63e91ad8748.	
6.1.1.2 **Reanalysis**	This component represents the software, processes and governance that provide *"a meteorological data assimilation project which aims to assimilate historical observational data spanning an extended period, using a single consistent assimilation*	Optional

	(or "analysis") scheme throughout", Wikipedia. See also NCAR at : https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables Examples of some available reanalysis include: • The European Centre for Medium-Range Weather Forecasts (ERA-5) • The Japanase Meteorological Agency (JRA-55). • The National Center for Atmospheric Research NCEP-NCAR (R1)	
6.1.1.3 Model Ensembles	This component represents the software, processes and governance that aggregate the data from: • a number of GCM and produce products that portray the range of model forecasts; • a number of runs of the same model	Optional
6.1.2 Generate Derived Data from Climate Observations		
6.1.2.1 Spatial Analysis	This component represents the software, processes and governance that undertake a very wide range of raster and vector spatial analysis techniques. Some examples are: • Generate grids that represent the spatial distribution of observations of an observed phenomena such as precipitation • Generate grids that represent distribution of average maximum temperature for the month of May for Climatological Standard Normals • Generate grids that represent distribution of maximum temperature anomalies for the month of May 2010	Recommended

	when compared to the Climatological Standard Normal • Select all meteorological stations that are within a buffer of ten kilometers around a national administrative boundary	
6.1.2.2 Image Analysis	This component represents the software, processes and governance that undertake a very wide range of image analysis techniques. Some examples are: • Process remotely sensed satellite imagery to determine the relative solar reflectance of a satellite image • Process remotely sensed satellite imagery to determine the cloud cover within a scene • Process remotely sensed satellite imagery to generate a Normalised Difference Vegetation Index to measure vegetation 'greenness' • Process ground based Radar imagery to detect rain and storm activity	Recommended
6.1.2.3 Time Series Analysis	This component represents the software, processes and governance that undertake the analysis of time series data using a very wide range of analysis techniques. Some examples are the analysis required to produce: • WMO Standard Products, e.g. Extremes, Standard Normals, World Weather Records, Climate Change Indices, etc A variety of Derived Observations Data. For more information see: • Guide to Climatological Practices (WMO-100), chapter 3 Statistical Climate Data Analysis, and Chapter 4 Climate Products	Recommended

6.1.2.4 Teleconnection Indices	This component represents the software, processes and governance to analyse, record and manage data representing Teleconnections and other major climate indices such as the El Nino Southern Oscilation (ENSO), or the Southern Ocilation Index (SOI). Wikipedia Teleconnection) (2023), defines: “**Teleconnection** in atmospheric science refers to climate anomalies being related to each other at large distances (typically thousands of kilometers).” For more information see: • Wikipedia https://en.wikipedia.org/wiki/Teleconnection (2023)	Optional
6.1.3 Data Homogenisation		
6.1.3.1 Data Homogenisation	This component refers to processes, software, governance and analysis of quality Observations data and ‘Observations Metadata’ to develop high quality ‘homogenised’ time series data sets. Such data sets aim to ensure that the only variability that remains in the time series is that resulting from actual climate variability. For more information see: • *Guidelines on Homogenization* (WMO-No. 1245) • The web site of the Joint CCl/Clivar/JCOMM Expert Team On Climate Change Detection And Indices (ETCCDI/CRD). See ETCCDI-1 (2020) • A current list of homogenisation packages may be found at http://www.climatol.eu/tt-hom/	Recommended
6.1.4 Disaggregation and infilling		

6.1.4.1 Disaggregation and infilling	For many applications in climate-sensitive activities, it is desirable to have complete time-series, with values available for each day. Therefore Members are encouraged to implement algorithms for infilling of missing data, and disaggregating accumulated totals of, say, rainfall into their constituent daily totals. As yet no standard WMO Guidelines are available on recommended practices, but these are expected to be developed in the future. In line with provenance requirements the original time-series data must be kept. It is strongly recommended that any estimated data due to infilling or disaggregation should be clearly flagged as such. This will depend on the quality flagging system in operation at the Member NMHS.	Recommended
6.1.5 Other		
6.1.5.1 Other	This component refers to as yet undefined processes, software, governance used to analyse climate data. Chapter 4 of the Guide to the Applications of Marine Climatology (WMO-No 781, 1994) provides some examples of the ways in which marine climate data is used to develop operational and monitoring products for marine and climate change applications.	Optional

6.2 References

6.2.1 WMO

Title	N°	Publication
Guide to the Applications of Marine Climatology	WMO-No. 781	2018

6.2.2 Links

Publisher	Title	URL	Current at
ET-CCDI	ET-CCDI Climate Change Indices	http://etccdi.pacificclimate.org/index.shtml	
ECMWF	ECMWF: Climate Reanalysis	https://www.ecmwf.int/en/research/climate-reanalysis	
WIKIPEDIA	Atmospheric Reanalysis	https://en.wikipedia.org/wiki/Atmospheric_reanalysis	
WIKIPEDIA	Teleconnection	https://en.wikipedia.org/wiki/Teleconnection	
Climatol	Task Team on HOMOGENISATION	http://www.climatol.eu/tt-hom/	
NCAR	Reanalysis	https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables	

7 Climate data presentation

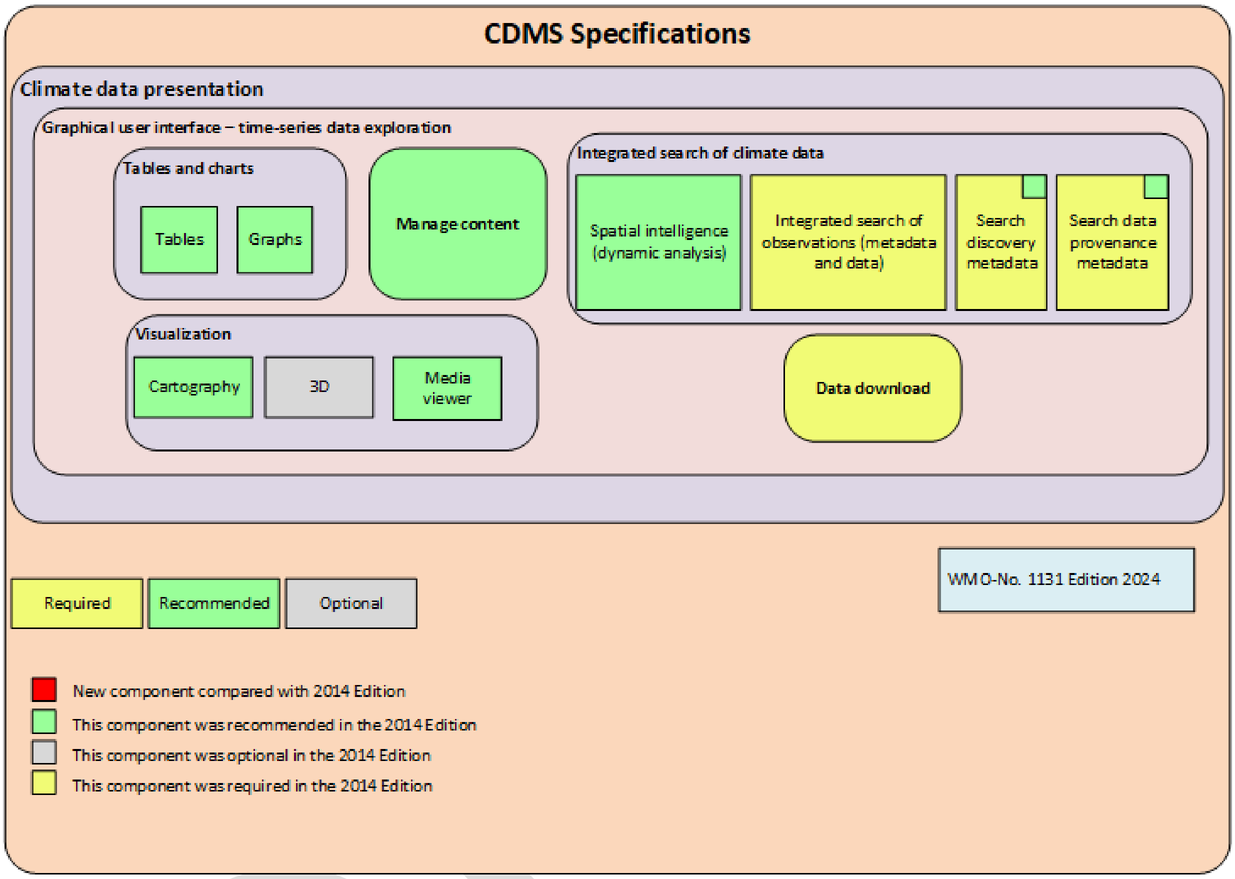


Figure 7.1. Climate data presentation

The Climate data presentation component represents the range of functionalities typically required to effectively visualize climate data.

7.1 Graphical user interface – time-series data exploration

7.1.1 Tables and charts

This subsection represents the technology, software, processes and governance mechanisms suitable for generating a broad array of tabular and graphical reports to effectively communicate displays relating to climate data.

7.1.1.1 Tables	This component refers to the technology, software, processes and governance processes suitable for generating a wide variety of tabular reports to effectively communicate issues relating to climate data. Examples include rainfall intensity, monthly summaries, etc.	Recommended
7.1.1.2 Graphs	This component concerns the technology, software, processes and governance processes suitable for generating a large variety of graphs to effectively convey climate data issues. Graphs could be presented in a wide array of formats, including: • Scatter plots • Histograms • Windroses • Time-series graphs using one or more variables For more information, see: *Guide to Climatological Practices* (WMO-No. 100), Chapter 4 Climate Products	Recommended
7.1.2 Manage content		
7.1.2.1 Manage content	This component covers the technology, software, processes and governance processes suitable for generating a wide variety of content to effectively communicate issues relating to climate data. This includes: • Developing products that integrate text, tables, graphics, different data sources e.g. monthly climate summary	Recommended

	• Preparing texts, documents and data for effective web presentation • Using technology such as content management systems to simplify web content presentation. See e.g. https://en.wikipedia.org/wiki/Content_management_system • Implementing effective governance processes that review, validate and authorize content prior to being published • Documentation/archive of content presentation	
7.1.3 Visualization		
7.1.3.1 Cartography	This component represents the technology, software, processes and governance processes suitable for generating a wide variety of cartographic output to effectively convey climate data products. It includes: • Spatial data preparation • Cartography • Simple point-and-click web maps • Symbology • Maps	Recommended
7.1.3.2 3D	This component provides the technology, software, processes and governance mechanisms suitable for visualizing and exploring climate data and issues within a 3D environment.	Optional

7.1.3.3 Media viewer	This component refers to the technology, software, processes and governance processes that enable various media to be displayed within the graphical user interface. Some examples are: • Photographs • Diagrams • Scanned documents such as scanned station records • Videos Recorded audio media	Recommended

7.1.4 Integrated search of climate data

7.1.4.1 Spatial intelligence	This component represents the technology, software, processes and governance processes that support an effective and dynamic analysis of climate data within a web environment to facilitate understanding of climate matters and communicate information based on or related to climate data. This dynamic analysis includes: • Geographical Information System (GIS) functionality, including the ability to perform spatial overlay analysis such as selecting points in a polygon. • The ability to search features by attribute, for example: ✓ Conducting a search of all stations within the catchment of a specific river ✓ Filtering the resultant stations to view only those that observe precipitation ✓ Viewing summary observations data for each of those stations This component integrates into a map-based interface a wide range of time-series data	Recommended

	including climate observations, climate grids, satellite imagery and topography, together with appropriate textual and other attribute data, such as climate metadata. It also facilitates dynamic data exploration and analysis using a broad array of integrated media, including maps, charts, graphs, tables and written reports.	
7.1.4.2 **Integrated search of observations (metadata and data)**	This component concerns the functionality that allows an end-user to conduct an integrated search of the climate database and the observations metadata catalogue. The search results will contain observations data and observations metadata. Some examples are: • Determining what observations data are available based on a set of parameters and viewing the results in a table. These parameters may include: ✓ Station ✓ Sensor or procedure ✓ Phenomena ✓ Data quality (based on quality flags, the climate observation quality classification or other method) ✓ Time period ✓ A variety of other observation metadata parameters • Reviewing observations metadata for selected stations. • Determining what datasets provide the actual observations data for a given station, sensor and phenomenon combination, together with the URL of the relevant discovery metadata records. The discovery metadata records will in turn provide the URLs of any services providing	Required

	dynamic access to the data. An example could involve searching for stations that use both a tipping bucket raingauge and manual methods to observe rainfall For more information, see: • Section on climate metadata (4.3) • Subsection on observations metadata (4.3.1) • Observations metadata catalogue component (8.2.1.2) • Linked data component (8.2.2.1)	
7.1.4.3 Search discovery metadata	This component refers to the functionality that allows an end-user to search the CDMS discovery metadata catalogue to: • Determine what datasets are managed by the NMHS. This search may be limited to datasets that are available publicly or those that are only available for internal use • Search for datasets in accordance with WIS parameters, categories and keywords • Review discovery metadata records to enable searchers to determine whether a dataset is suitable for their particular use • Determine the URL that can be used to access online services that host the dataset for dynamic access and data download This same component could be used to search WIS metadata catalogues, and thereby should be taken to include discovery metadata requirements into any datasets made available for Regional and global analysis For more information, see: • Section on climate metadata (4.3)	Required

	• Subsection on dataset discovery metadata (4.3.2) • Discovery metadata catalogue component (8.2.1.1) • Linked data component (8.2.2.1)	
7.1.4.4 **Search data provenance metadata**	This component provides the functionality that allows an end-user to search the CDMS data provenance metadata catalogue to: Broadly determine the lineage of a dataset, including the processes the dataset has been subjected to. Trace the provenance of individual records in detail, taking into account: • What was changed? • What was it derived from? • When was it changed? • What was done to change it? • How and why was it changed? • Who changed it? • Who did they act on behalf of (if applicable)? • Who authorized the change? For more information, see: • Data lineage traceability (3.1.6.2) • Section on climate metadata (4.3) • Subsection on data provenance (4.3.3) • Data provenance metadata catalogue component (8.2.1.3) • Linked data component (8.2.2.1)	Required
7.1.5 Data download		

7.1.5.1 Data download	This component represents the functionality enabling end-users to download climate data. This component is related to the climate data delivery components (Chapter 8) and data discovery registers.	Required

8 Climate data delivery services

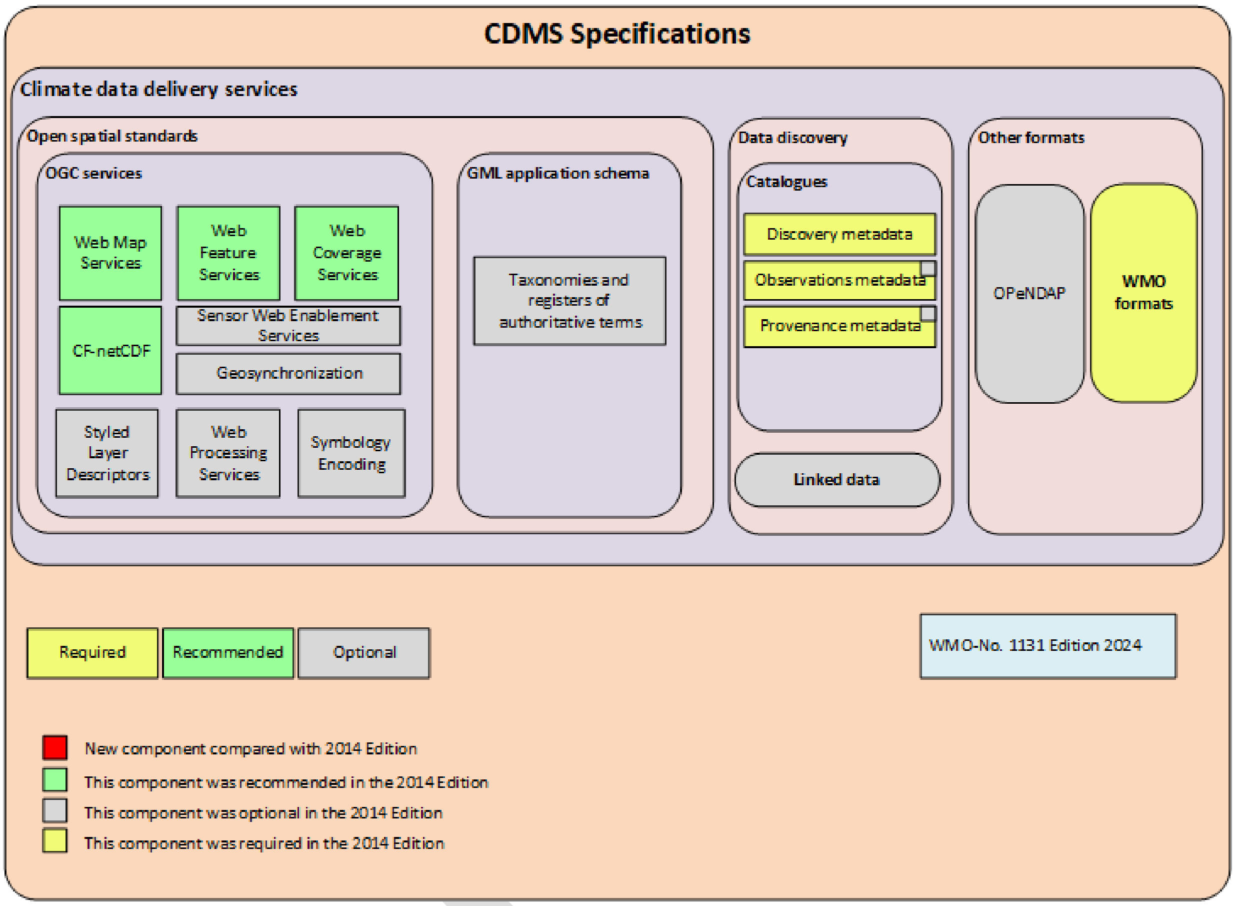


Figure 8.1. Climate data delivery services

The Climate data delivery services component represents the range of functionalities typically required to effectively deliver climate data, both internally and externally.

8.1 Open spatial standards

This section has been included because open spatial standards are seen as a mechanism that is being increasingly adopted by many organizations and industry sectors around the world. These types of services underpin global attempts at making data easily accessible through initiatives such as the Global Earth Observation System of Systems (GEOSS) (see GEOSS web page). The WMO Information System can be considered as a component of GEOSS.

Open spatial standards are being increasingly supported by a wide range of off-the-shelf software, including traditional desktop GIS software, making it easier for users to access data that are served via such services.

This is particularly important for CDMSs, as there is a large potential user base that does not routinely use climate data but could benefit from having a reliable source from which to take the data to ensure consistent use across industry. Some examples of growing interest in climate data can be found in sectors such as agriculture, emergency services, aquaculture, fishing, tourism, transportation, health and environment. Such industries typically do not understand WMO data formats such as BUFR or GRIB, nor do they understand how to exploit data delivered in such formats.

While developing countries may not have reliable access to the Internet to take advantage of external open spatial services, it would certainly be possible to use them within their own internal local area networks, particularly as a means to visualize their data.

In short, open spatial standards are expected to become an increasingly important mechanism for distributing data in future years.

Note: The open spatial components presented below are indicative of the types of standards and services that are available and appropriate for the delivery of climate data. These components are not intended to be exhaustive as there are many more services and standards that are also relevant. It is anticipated that this will be expanded upon in future revisions of this publication.

For more information, see:

- The OGC Abstract Specifications, which provide a detailed theoretical overview of the theory and rationale underpinning open spatial standards.
- The ISO/Technical Committee 211 Advisory Group on Outreach *Standards Guide: ISO/TC 211 Geographic Information/Geomatics*.

8.1.1 Open Geospatial Consortium services

8.1.1.1 Web Map Services	This component represents technology suitable for the distribution of a wide range of climate data via a Web Map Service (WMS). In essence, a WMS provides a map view of data distributed via a georeferenced image. For more information, see: OGC WMS documentation	Recommended

	It includes the OGC-API standards.	
8.1.1.2 Web Feature Services	This component represents technology suitable for the distribution of a broad range of vector climate data via a Web Feature Service (WFS). In essence, a WFS could provide vector and tabular climate data, which could be presented in a number of formats such as Geography Markup Language(see OGC GML web page) or Environmental Systems Research Institute (ESRI) shapefile. Some WFS implementations can serve data constrained by a logical data model (also known as an application schema). It may also be possible to enable WFS server software to provide meteorological observation data via WMO formats such as BUFR. There may be issues with serving time series via WFS. There is discussion that a Sensor Web Service may be better for time-series observations. For more information, see:OGC WFS documentation	Recommended
8.1.1.3 Web Coverage Services	This component represents technology suitable for the distribution of a wide range of gridded and array climate data via a Web Coverage Service (WCS). In essence, a WCS provides the actual gridded or array data. Future versions of WCSs are intended to support logical data models defined as application schemas. For more information, see: OGC WCS documentation	Recommended

8.1.1.4 Sensor Web Enablement Services	This component represents a range of technological tools suitable for the dissemination (and collection?) of a wide variety of observational data and related metadata. Sensor Web Services are typically used with data that conform to the observations and measurements data model. This model is well suited to time-series data. Therefore, Sensor Web Services may be an appropriate mechanism for serving time-series climate data in the future. For more information, see the OGC documentation for: • [Observations and measurements] • [Sensor Observation Services]	Optional
8.1.1.5 CF-netCDF	This component involves technology suitable for the provision of a wide variety of gridded or array scientific data written as netCDF files that support the conventions for climate and forecast metadata[i](In this context, the term metadata refers to a set of fields in the header of a netCDF file that describe the context and format of the array data contained in the CF-netCDF file). For more information, see: • Wikipedia article on netCDF • OGC netCDF standards suite	Recommended
8.1.1.6 Geosynchronization	This component concerns a series of technological tools that enable a data publisher to distribute a data product in an environment that supports managed change to the source data to retain consistency. In theory, end-users could subscribe to a data source and have changes to the data replicated in their copy of the data.	Optional

	In summary, geosynchronization services are expected to support several processes, including: • Allowing interested parties to subscribe to an authoritative data source. • Data entry with validation. • Notifying interested parties of changes. • Allowing replication of the data provider's features. It is anticipated that geosynchronization services will support crowdsourced data and processes in addition to authoritative data sources. At present, geosynchronization services are being developed with the current version of a standard that only serves data via a WFS. It is expected that geosynchronization will support additional services in the future, including WCS and WMS. For more information, see: • OGC overview of geosynchronization services • The OWS 7 engineering report on the test of geosynchronization services	
8.1.1.7 Web Processing Services	This component covers a range of technological instruments that provide a standards-based framework for developing spatial processing services that operate via an internal network or the Internet. This standard is being used by a number of open-source projects and vendors to develop the building blocks that will support a wide range of spatial analytic processes. The latest version of the standard is being developed to support both synchronous and	Optional

	asynchronous Web Processing Services (WPS). This standard has considerable potential for future CDMS use, for example: To enable an NMHS to establish a suite of services to process and analyse climate data within a services-oriented architecture. For example creating subsets from a large dataset. To enable a future service-provider to offer CDMS-related services via “the cloud”. For more information, see: • WPS Wikipedia entry • OGC WPS documentation	
8.1.1.8 Symbology Encoding	This component represents a range of technological tools that provide rules and a standardized approach for defining alternative visual portrayals of spatial data via an internal network or the Internet. Symbology Encoding, together with Styled Layer Descriptors (SLDs), can be used with WMSs, WFSs and WCSs to enable user-defined symbolization of spatial data from within a collection of published styles. As an example, it is possible to publish several colour classification schemes for a gridded dataset, allowing end-users to select one that is appropriate for their use. For more information, see: • OGC Symbology Encoding documentation	Optional
8.1.1.9 Styled Layer Descriptors	This component represents a range of technological tools that provide rules and a standardized approach for defining alternative visual portrayals of the spatial data via an internal network or the Internet.	Optional

	Styled Layer Descriptors, together with Symbology Encoding, can be used with WMSs, WFSs and WCSs to enable user-defined symbolization of spatial data from within a collection of published styles. For more information, see: • OGC SLD documentation	

8.1.2 Geography Markup Language application schema

Application schemas (as defined in ISO 19109:2005 *Geographic information – Rules for application schema*) provide an abstract representation of the content and structure of information resources. The Climate observations application schema component (4.2.3.2) outlines the use of application schemas specifically developed for the climate domain.

These abstract representations of information provide the basis for deriving concrete encodings or data formats that allow the information to be serialized for exchange between systems and organizations.

In addition, the application schema can be used to develop XML-based data encodings using widely supported open standards for geographic information. The ISO 19136:2007 *Geographic information – Geography Markup Language (GML)* standard provides rules for serializing the abstract data model expressed as an application schema via XML encoding to create a GML application schema.

In summary:

- An application schema can be thought of as a logical data model. For more information, see the subsection on WMO logical data models (4.2.3).
- A GML application schema is a physical model that is derived from a logical data model published using a particular technology – which in this case is GML.

Deriving a GML application schema from an application schema developed specifically for the climate domain (see Climate observations application schema component (4.2.3.2)) is anticipated to make climate data far more readily consumable for a broader community of users such as those interested in determining the impacts of climate change.

8.1.2.1 Taxonomies and registers of	This component represents the technology, software, processes and governance needed to develop an authoritative definition of the concepts and terms referenced in a logical	Optional

authoritative terms	data model and to enable the publication of such terms.	

1

8.2 Data discovery

8.2.1 Climate metadata catalogues

8.2.1.1 Discovery metadata catalogue	This component refers to the technology and processes that create a discovery metadata catalogue. This catalogue is used to publish an organization's data holdings as discovery metadata records, with corresponding records describing which online services may be used to access each dataset, or whether restrictions apply. A discovery metadata catalogue allows an organization to participate within the WIS environment.	Required
8.2.1.2 Observations metadata catalogue	This component refers to the technology and processes that create the observations metadata catalogue used to publish an organization's observations metadata records. It is anticipated that the climate database will be used to store and manage observations metadata. This component will serve as an IT catalogue service for observations metadata. More work is required to define this component.	Required
8.2.1.3 Data provenance	This component refers to the technology and processes that create the data provenance metadata catalogue used to publish an	Required

metadata catalogue	organization's data provenance metadata records. A compliant CDMS should store and manage data provenance metadata, or at least be linked to another database that does this. This component will serve as an IT catalogue service for data provenance metadata.	

8.2.2 Linked data

8.2.2.1 Linked data	This component supports semantic search requests such as those used with linked data. This is an emerging requirement that is building considerable momentum within information management communities. The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset, published by the Australian government at https://dl.acm.org/doi/abs/10.3233/SW-160241 provides an example of how linked data may be used for publishing climate data. More work is required to define this component, including its relationship to the approaches adopted by WIS. For more information, see:Linked data	Optional

8.3 Other formats

8.3.1 WMO Formats

	This component represents technology suitable for the distribution of a wide range of climate data via traditional WMO formats. The technology might be available within the CDMS itself, or otherwise should be accessible via WIS. For historical data, however, knowledge of cancelled or already cancelled formats will always be needed. For more information, see: • FM 94 BUFR Edition 4 • FM 92 GRIB Edition 2 Other formats also exist, but it is anticipated that they will be phased out in time. For more information, see: • *Manual on Codes* (WMO-No. 306), Volume I.2 • Wikipedia articles on BUFR and GRIB	Required

8.4 References

8.4.1 WMO

Title	#	Publication
Manual on Codes	WMO-No. 306	2019

8.4.2 Links

Publisher	Title	URL	Current at
GEOSS	GEOSS Portal	https://www.geoportal.org/?m:activeLayerTileId=osm&f:dataSource=dab	
ISO	ISO 19156:2011 Geographic Information – Observations and Measurements	https://www.iso.org/standard/32574.html	
ISO	ISO/TC-211 Geographic Information/Geomatics	https://www.iso.org/committee/54904.html	

NASA	The Data Access Protocol -	https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/data-access-protocol-2
OGC	OGC Geographic Markup Language Web page	https://www.ogc.org/standard/gml/
OGC	Abstract specification	https://www.ogc.org/standards/as/
OGC	OGC Web Map Services	https://www.ogc.org/standard/wms/
OGC	OGC Web Feature Service	https://www.ogc.org/standard/wfs/
OGC	OGC Web Coverage Service	https://www.ogc.org/standard/wcs/
OGC	OGC Sensor Obervations Service	https://www.ogc.org/standard/sos/
OGC	OGC Network Common Data Form (NetCDF) standards suite	https://www.ogc.org/standard/netcdf/
OGC	OGC Standards Working Groups	https://www.ogc.org/about-ogc/committees/swg/
OGC	OGC Web Processing Service	https://www.ogc.org/standard/wps/
OGC	OGC Symbology encoding	https://www.ogc.org/standard/se/
OGC	OGC Geoserver Styled layer descriptor	https://docs.geoserver.org/stable/en/user/styling/sld/reference/index.html
OGC	OWS 7 Engineering Report -- Geosynchronization service OGC 10-069r2	file:///C:/Users/wjwri/Downloads/10-069r2_OWS_7_Engineering_Report_--_GeoSynchronization_service.pdf
Wikipedia	NetCDF	https://en.wikipedia.org/wiki/NetCDF
Wikipedia	Web Processing Service	https://en.wikipedia.org/wiki/Web_Processing_Service
Wikipedia	Linked data	https://en.wikipedia.org/wiki/Linked_data
Wikipedia	OPeNDAP	https://en.wikipedia.org/wiki/OPeNDAP
Wikipedia	BUFR	https://en.wikipedia.org/wiki/BUFR
Wikipedia	GRIB	https://en.wikipedia.org/wiki/GRIB

1

2

9 Core IT infrastructure

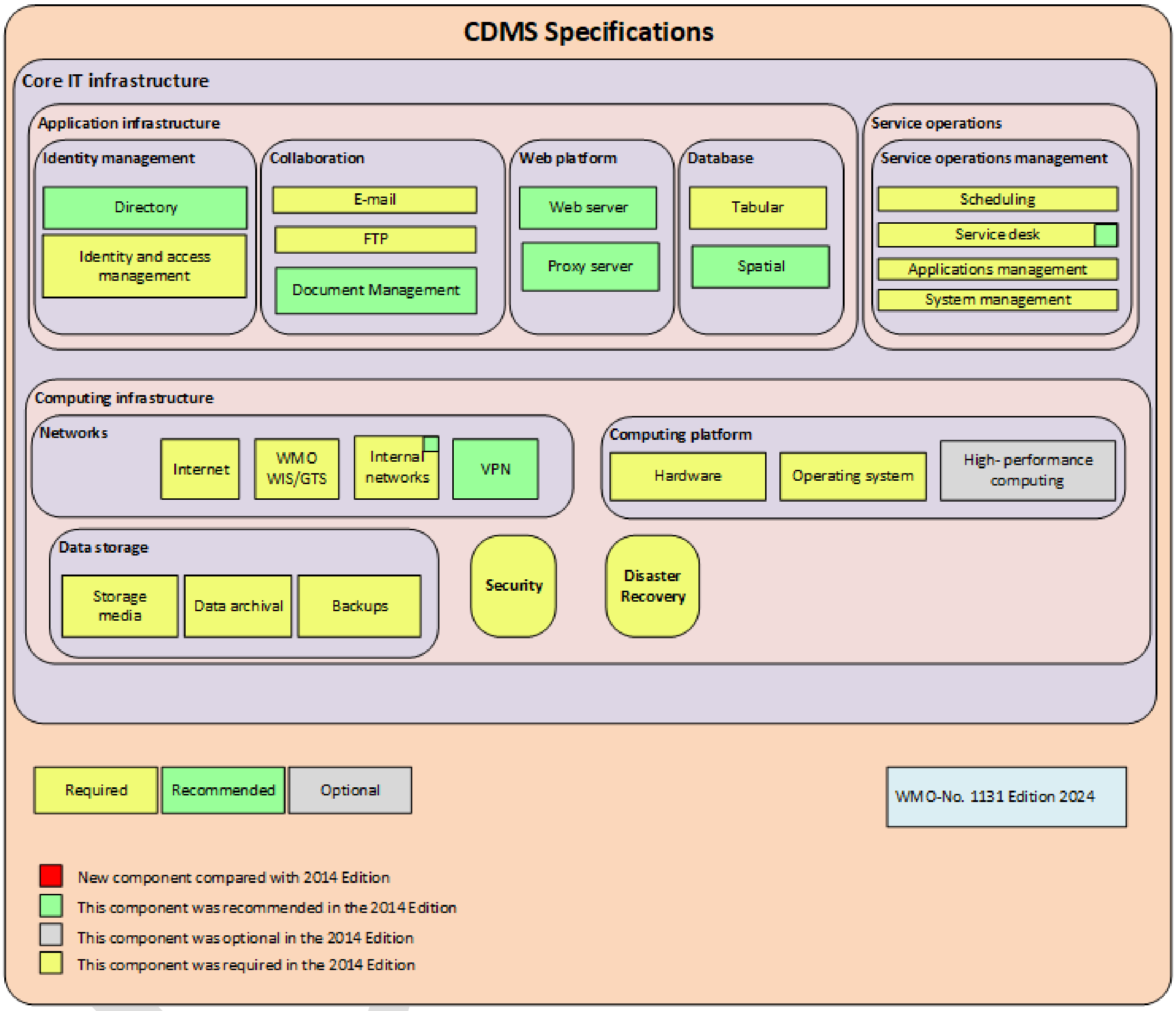


Figure 9.1. Core IT infrastructure

The Core IT infrastructure component represents the underlying IT functionalities typically required to support a CDMS.

9.1 Application infrastructure

9.1.1 Identity management

9.1.1.1 Directory	This component provides directory services such as the Lightweight Directory Access Protocol or Active Directory to manage user credentials and details.	Recommended
9.1.1.2 Identity and access management	This component supports policies and functionalities that enable granular user access to the organization's IT resources and data.	Required

9.1.2 Collaboration

9.1.2.1 E-mail	This component provides secure e-mail access and includes functionalities such as filtering for malware and spam. In this context it refers specifically to data management and related IT support functionality, but should be part of a whole-of-NMHS security/policy infrastructure.	Required
9.1.2.2 FTP	This component provides secure services to allow exchange of climate data via the use of the (Secure) File Transfer Protocol ((S)FTP).	Required
9.1.2.3 Document management	This component supports a collaborative web environment allowing any member of a team to easily edit and publish content (Wiki, document management platform, etc.).	Recommended

9.1.3 Web platform

9.1.3.1 Web server	This component provides functionalities that deliver web content to web browsers. In addition to the web server platform, it also refers to services required to support web applications.	Recommended
9.1.3.2 Proxy server	This component routes web traffic and acts as a load balancer and a reverse proxy server to contribute to secure connections to the web server.	Recommended

9.1.4 Database

9.1.4.1 Tabular	This component represents database technology suitable for the storage of a wide range of time-series climate data in tabular format, typically within a relational database.	Required
9.1.4.2 Spatial	This component deals with technology used to spatially enable time-series climate data, typically within a relational database. The component may consist of a functionality that spatially enables the tabular database component, or it could be a dedicated spatial database that is closely aligned to the climate data stored within the tabular database.	Recommended

9.2 Service operations

9.2.1 Service operations management

9.2.1.1 Scheduling	This component represents technology and processes used to ensure that software processes can be scheduled to run at specific times e.g. every 24 hours, every weeks, every month, etc. This functionality supports activities such as regular data ingest, quality assurance operations, data analysis, derivation and backups. There is also a requirement in more advanced environments to ensure that the dependencies between scheduled jobs are managed.	Required
9.2.1.2 Service desk	This component represents the functionalities, processes and software required to provide support for service operations, including: Incident and event management to ensure that if an unplanned interruption to an IT service occurs, normal service operation is returned as soon as possible. Problem management to ensure that the root causes of problems are found and where possible, rectified. See: • Wikipedia on Information Technology Infrastructure Library (ITIL): https://en.wikipedia.org/wiki/ITIL • ITIL Service Operation that highlights best-practice for delivering IT to meet agreed service levels for both business users and customers.	Required

9.2.1.3 **Applications management**	This component covers the functionalities, processes and software required to provide application administration tasks and second- and third-level support for CDMS services. Any new IT system implementation and any change to existing IT systems must be undertaken in accordance with the section on governance (3.2).	Required
9.2.1.4 **Systems management**	This component refers to the functionalities, processes and software required to provide systems administration tasks and second- and third-level support for CDMS services. This might include updates and upgrades to software, patches, security etc. Any new IT system implementation and any change to existing IT systems must be undertaken in accordance with the section on governance (3.2).	Required

1

9.3 Computing infrastructure

9.3.1 Networks

9.3.1.1 **Internet**	This component covers the infrastructure required to support access to the Internet. This includes routers, switches, firewalls, internet service providers, etc.	Required
9.3.1.2 **WMO WIS/GTS**	This component concerns the infrastructure required to access the WMO network (WIS, GTS). This is essentially a private wide-area network.	Required

9.3.1.3 Internal networks	This component refers to the infrastructure required to support local area networks. This includes switches, firewall(s), services such as domain name servers or the Dynamic Host Configuration Protocol.	Required
9.3.1.4 VPN	This component concerns a virtual private network (VPN), which allows a private network to be set up across the publicly available Internet making use of tunnelling and security features. This can result in relatively secure communications.	Recommended

9.3.2 Computing platform

NMHSs now have several choices about their computing platforms. Rather than run all their computing operations within the NMHS, the possibility exists to opt for a Cloud solution (public or private), or a hybrid Cloud solution

9.3.2.1 Hardware	This component covers all computing hardware, including servers and desktop computers. Organizations are increasingly using virtualization to allow several virtual servers to be deployed on a single physical server, as a way of minimizing hardware and operational costs while increasing operational efficiency.	Required
9.3.2.2 Operating system	This component concerns the operating system required to support computing operations.	Required
9.3.2.3 High-performance computing	This component covers the advanced computing functionalities needed to support high-performance computing, including clusters and grids.	Optional

9.3.3 Security

9.3.3.1 Security	Security is actually an aspect of all components, but is included here for clarity. All software and systems should be implemented with security in mind in order to protect the integrity of climate-related systems and data. This component does not just refer to IT security but also to physical security, such as preventing the theft of a server and the resulting loss of the climate database. It includes provisions for enforcing the governance/authentication processes referred to above, procedures for identifying/tracking unauthorized access, security upgrades, etc.	Required

9.3.4 Data storage

9.3.4.1 Storage media	This component involves the provision of sufficient storage media to cover operational activities, including the storage of climate data, systems, archives, backups and disaster recovery materials.	Required
9.3.4.2 Data archival	This component handles the secure archival of historical data to ensure that it is available for future generations.	Required

9.3.4.3 Backups	This component covers the regular operational backup and restoration of data and systems. It includes both regular on-site back-ups, and off-site back-ups.	Required
9.3.5 Disaster recovery		
9.3.5.1 Disaster recovery	This component refers to the disaster recovery and business continuity policies, processes, plans and systems required to ensure that CDMS systems and climate data can be recovered in the event of an unforeseen incident. This could be as simple as a server malfunction or as complex as an organization's city being destroyed due to some unexpected event such as an earthquake, tsunami or military action. This is why off-site storage of backups is required, along with regularly-tested procedures to restore functionality as quickly as possible.	Required

1

10 Considerations

There are many aspects that organizations are advised to consider when implementing a new CDMS. As each organization will have different requirements, this chapter will only suggest and outline sufficient issues to provide organizations with a starting point for undertaking their own planning.

These issues are provided as a guide only. Organizations will need to decide what is relevant for their particular purposes. However it must be emphasized that a new CDMS should be designed in such a way as to incorporate relevant WMO specifications.

The issues discussed briefly in this chapter are:

- Where to start?
- How much can the organization afford to spend on implementing the CDMS and maintaining it on an ongoing basis?
- What is the current state of the management of climate data and related systems?
- What functionality would the organization like to implement?
- What policies are in place to ensure that the incorporation of new data types are sustainable and affordable?
- What CDMS options are there?
- How can existing data and systems be transitioned to the new CDMS and how can staff be assisted in the transition?
- What skills and resources will be required to implement and support the new or revised CDMS?
- What are the next steps?

This thorough assessment is just as crucial for developing countries as for developed countries.

It is important to ensure the long-term viability of both the CDMS and the climate record. A change to a CDMS could lead to very serious unforeseen consequences for the climate record. Care needs to be taken during the planning stage to try to mitigate this risk as much as possible.

Failure to exercise appropriate due diligence increases the risk of a failed CDMS implementation that may be very expensive to resolve.

Note: It is understood that different organizations will have different procurement processes that they are required to follow when acquiring new products and services. This publication will not discuss procurement issues, as organizations will be in the best position to assess procurement implications for themselves.

10.1 Where to start?

Establish a small team to conduct careful planning work as is suggested below. The team may decide that additional effort is required and may consider other relevant items of interest. The team should take into account the issues described below, and many of the issues described in Section 10.3; see Section 10.8.2 for more details.

It is suggested that the team:

- Have a clearly understood mandate agreed and adequately financed by the NMHS Executive to perform the work and a commitment to fund maintenance of the system.
- Clearly define the business problem the team is trying to resolve in preparation for the later development of a business case.
- Undertake appropriate analysis to understand how the issues discussed in this chapter affect the organization.
- Ensure that the results of the analysis are documented and taken into consideration and that an effective implementation plan, project plan and data management plan are developed.
- Thoroughly consider all aspects of governance, ingest, quality management, data access, as outlined in earlier Chapters.
- Consider aspects such as staff training, data migration, documentation, etc, and how to avoid disruptions to routine operations while these important tasks are undertaken.
- Be comprised of experienced senior staff (and possibly appropriate consultants) with a broad mix of skills. One person may well have skills in several disciplines. Suggested skills required for this work include:
    - ✓ Climate science
    - ✓ Climate data management
    - ✓ IT architecture
    - ✓ Project management
    - ✓ IT service management
    - ✓ Software development
    - ✓ IT systems and administration
    - ✓ Climate service delivery
    - ✓ Spatial Data, Spatial Systems, Spatial Analysis

10.2 Funding

The team should determine the level of funding available to the organization for:

- Undertaking project tasks to implement a CDMS.
- Running the day-to-day service and data management activities needed to support an operational CDMS.

It would also be beneficial to determine whether any additional funding sources exist that could support CDMS implementation and management. This might apply, for instance, where a developing country is supported by bilateral or multilateral funding arrangements; however there needs to be some mechanism to ensure that any such arrangements provide for ongoing sustainability (see the publication "*A Proposed Long-Term Resource Plan for Climate Data Management and Rescue activities*" (WMO, 2017).

In addition, it would facilitate the planning process if team members could identify potential CDMS implementation costs during planning activities. Such costs might include, for instance, system licensing costs (if applicable), data migration costs, development of new functionality (if required), system upgrade and replacement costs, and other costs associated with issues as outlined in Section 10.3 below.

This information will provide the team and the organization's management with suitable background information to establish the scope and duration of the CDMS implementation at an affordable level.

10.3 Current state

The team should ideally develop an understanding of the organization's ability to meet CDMS requirements in their current environment.

Some suggested areas to investigate are:

- Service management
  - ✓ What service-level agreements exist?
  - ✓ What value do the services provide to the NMHS, and its users?
  - ✓ How is that value measured and monitored?

- Business requirements
  - ✓ What business need is being met by current CDMS processes? For instance, supporting NMHS, national users, and broader climate service requirements via the acquisition stewardship and provision of climate data.
  - ✓ Which of the CDMS components described in this publication are being met by current approaches?
  - ✓ What CDMS governance processes exist and do not exist?

- Data issues
  - ✓ What data holdings are being managed?
  - ✓ What is the format of the data?
  - ✓ Where are the data currently located?
  - ✓ What data flows exist within the organization?
  - ✓ How much disk space is required to store the data?
  - ✓ What is the anticipated annual growth rate in data storage requirements?

- ✓ What backup processes are in place and where are the backup copies stored?
- ✓ What disaster recovery (or business continuance) processes are in place to protect the integrity of the climate record and operational systems?
- ✓ Are there climate metadata records (discovery, observations and provenance) that describe the climate data?
- ✓ What quality assurance processes and flags are used in relation to observations data?
- ✓ What data license and intellectual property issues relate to the data?
- ✓ Are any third-party data being managed within the archive?
- ✓ What other outstanding data issues are known?

- Application software issues
  - ✓ What application software is being used?
  - ✓ What software licenses apply to the software?
  - ✓ Are there any annual support fees associated with the software?
  - ✓ Where is the software source code?
  - ✓ Where is the software deployed?
  - ✓ Who is using the software within the organization and elsewhere and could thus be impacted by a CDMS change?
  - ✓ What scheduled tasks are being run?
  - ✓ What implications will the decommissioning of the software have on business need?
  - ✓ Which staff members maintain the software? Noting that inevitable staff turnover requires an allocation of specific duties to a verified position.
  - ✓ A generic need to ensure adequate “handover” training in the event of staff turnover.
  - ✓ Where are the software deployment instructions?
  - ✓ Are there any component configuration databases?

- IT infrastructure issues
  - ✓ What IT infrastructure is being used?
  - ✓ What is the life expectancy of the IT infrastructure?
  - ✓ Does it need to be replaced?
  - ✓ What operating system is being used?
  - ✓ Which staff members maintain the IT infrastructure?
  - ✓ What level of capacity does the IT infrastructure have to support a new CDMS?
  - ✓ Does Cloud hosting, meaning data, IT Infrastructure hosting and Software, as a Service, provide an alternative, and what policies and procedures are required to ensure that any data hosted externally are secure and recoverable?

10.4 New functionality

The team should identify and examine the new CDMS requirements and functionalities they would like to implement in order to meet their business needs.

Service designers should have a clear and realistic understanding of the relative priorities of each of the requirements selected, as funding and resource considerations may dictate what can be implemented and how long the implementation can take.

10.4.1 Determine functional requirements

Determine the functional requirements to be implemented in the new or revised CDMS, together with the relative priorities of each requirement.

When conducting this work, the team should consider:

- Using this CDMS Specifications as a guide. The CDMS component classification scheme may be useful in determining relative priorities for the implementation of each functionality.
- The functionality and capability that already exist within the organization (determined during the assessment of the current state of the CDMS).
- The amount of project funding available, and the funding available for ongoing support.
- Whether the functional requirements can better be provided utilizing the WIS, and/or functionality of Regional Climate Centres.

10.4.2 Determine non-functional requirements

In addition to the functional requirements, a number of non-functional requirements relating to the operation of the CDMS will also need to be considered. According to the Wikipedia article (https://en.wikipedia.org/wiki/Non-functional_requirement), "*In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. They are contrasted with functional requirements that define specific behaviour or functions*".

It illustrates this difference by way of an example: "A system may be required to present the user with a display of the number of records in a database. This is a functional requirement. How current this number needs to be, is a non-functional requirement. If the number needs to be updated in real time, the system architects must ensure that the system is capable of displaying the record count within an acceptably short interval of the number of records changing."

Some suggested issues to consider are:

- Ensuring that suitable policies and governance are in place to support the day-to-day operations of the CDMS. See Chapter 3 of this publication.

- Evaluating the existing backup and disaster recovery procedures in place and revising as appropriate to ensure the long-term viability of the CDMS. Note: these should regularly be tested.
- Ensuring sufficient redundancy in data and systems to meet operational requirements.
- Ensuring that the IT infrastructure has sufficient capacity to support the expected demand of the new CDMS.
- Evaluating data and system portability. It is possible that over the life span of the CDMS, it may need to be ported to a new software or hardware platform. Therefore, strong consideration should be given to mandating open systems and open data.
- Taking account of the probable need to scale the implemented solution as demand increases and, therefore, being mindful of potential upgrade paths when designing various components.
- Examining the need to support the internationalization of components. For instance, Is there an anticipated need to support more than one language within the user interface and output products?
- What options are available to support various components?
- Taking account of the findings of the UK parliamentary inquiry into Climategate (UK Parliament Science and Technology Committee, 2011) and the need to store, and if necessary make available, the software source code that has been used to analyse, process and transform the climate data. This implies the need for a solution that uses open-source software. Note, however, that while the souce code for the analysis, paramaterisations, and derivations of data should be treated with this transparency, the software that controls the processing pipeline does not need this same level of transparency, and may even be commercially-available off the shelf software".
- Taking into account the implications of implementing a CDMS for IT service management. For example:
  - ✓ What service is to be offered?.
  - ✓ How is the service to be defined
  - ✓ What is the value of the proposed service for the organization
  - ✓ What key performance indicators will be implemented to assess the performance of the service and how will they be measured
  - ✓ What factors will determine a successful implementation? How will they be measured?

10.5 Evaluate options

Once the team has a clear idea of the functionality they desire, they can evaluate the options available for delivering that functionality. Again, this CDMS Specification may be useful in evaluating a product.

A number of options are available to organizations, including:

- Implementing open-source or proprietary software that meets the functionality of one or more components of the CDMS. All required functionalities do not have to be delivered within one product. Many successful proprietary and open-source products will meet the functionality of several components.
- Implementing an existing CDMS, taking into account:
  - ✓ How well does the functionality offered by the CDMS align with the requirements of this CDMS Specification?
  - ✓ There are a number of CDMSs that are currently available. However, while many systems call themselves a CDMS, there was no consistent definition of the functionality expected within a CDMS until this publication was prepared. The available functionalities will therefore vary within each product.
  - ✓ What level of support is available for the CDMS?
  - ✓ How is the CDMS customized and extended?
  - ✓ Is the product open-source?
  - ✓ What resources are available to the organization that maintains the CDMS product?
- Implementing custom-developed CDMS functionalities.
- If the CDMS implementation is undertaken as part of a bilateral/multi-lateral aid project, what support is available should an NMHS need to add additional functionality?
- Under a suitable agreement, porting specific desired functionality from an existing CDMS (code-sharing libraries may assist in this regard);
- A combination of all of the above.

Regardless of the implementation approach taken, this CDMS Specification can be used to provide guidance on the type of functionality required in the CDMS.

At this stage of the process, organizations may decide to evaluate a number of options so that they have a clear understanding of the best way forward for their specific circumstances.

10.6 Examine transition issues

Once the team has decided on a shortlist of possible solutions, they should investigate a number of transition issues to develop a better understanding of the viability of each of the selected options.

Some example issues are discussed below.

10.6.1 Existing climate data

An organization will typically have a considerable amount of historical climate data that it maintains as part of its day-to-day activities. These data may be in digital form in an existing climate database, in spreadsheets, on tape or microfiche, etc., or in a hard copy format. A decision will need to be made as to whether the organization can afford to keep the data in their current form or whether they should be migrated to the new CDMS. Some issues to consider are:

- How can existing hard copy climate data be digitized?
- Is at least some existing climate data stored in formats that might soon be obsolete, or on media that are decaying? If so, it is imperative to design and implement a transition program to newer formats, preferably within the new CDMS.
- Can the organization afford to maintain the existing CDMS in addition to the new CDMS, both in terms of IT and staff skills, and for how long?
- What is the difference in the underlying data models between the current and new CDMS for both climate data and climate metadata?
- Can the data in the old CDMS be migrated without loss of data, a reduction in the integrity of the data or loss in the context of the data?
- If not, is it possible to undertake some transitionary work to retain as much of the data and context as possible?
- Do the old and new CDMSs apply the same approach to quality assurance and quality assurance flags? Is it possible to retain the context of previous quality assurance work?
- What is the best way to migrate climate data and metadata from the old CDMS to the new CDMS?
- What testing strategy is required to ensure the integrity of the organization's climate data and metadata once they have been successfully migrated to the new CDMS?
- What does the organization do with data that could not be migrated successfully to the new CDMS?
- If not already catered for, can climate data and associated observations metadata be better linked?

10.6.2 Deployment in stages

Rather than implementing a new CDMS as part of a monolithic deployment process, is it possible to break the implementation down into smaller stages, so that the NMHS can receive early benefits from the new system? If this can be done, it will decrease the risks associated with the deployment and allow a gradual transition to the new functionality.

What dependencies are there between each stage of the deployment?

What does a successful CDMS deployment look like? How can this be measured objectively?

10.6.3 Decommission redundant components

Once the new CDMS has been deployed and is operating successfully within a stable environment, what work needs to be done to decommission the old CDMS components? Note that adequate communication with service providers within the NMHS must be provided.

10.6.4 Training and Operating procedures

What training is required to ensure that staff can effectively deploy, manage, maintain and use the new CDMS? This includes not just technical maintenance of the hardware, but possibly the development of new skills, such as SQL training when transitioning to a relational database.

When should this training be provided to ensure that staff can capitalize on their training?

It is often worth planning several training sessions spaced out over time and according to software complexity. This would allow users to master the basic skills needed to use the CDMS and, once they are more familiar with the system and day-to-day work, to take more advanced courses.

In the case of CDMS deployment as part of an externally-funded aid project, it will be necessary to provide for refresher training, specifically including training in the operational environment of the NMHS(s) where the CDMS implementation is taking place, as outlined in the WMO publication “A Proposed Long Term Resource Plan for Climate Data Management and Rescue Activities” (WMO, 2017).

Alongside the training, operational procedures describing the users' day-to-day work with the CDMS must be written. This is the essential link between the CDMS functionalities and how users will use them to meet NMHS obligations in their country.

10.7 Skills and resources

It would be very useful if the team could assess the skills and resources required to implement, maintain and operate the new CDMS over the long term, as well as a suitable staff development/recruitment program to ensure these requirements are met in an ongoing way.

It is important that this be determined during the planning stage, so that a business case can be developed to ensure the long-term sustainability of the CDMS.

10.7.1 Skills

Consider the following mix of skills needed to operate and maintain the CDMS:

- Climatology;
- Data management;
- Statistics;
- Meteorology;
- Observations instruments and processes;

- Spatial information;
- IT architecture;
- Project management;
- Software development;
- IT service management;
- IT systems and administration
- Training.

10.7.2 Determine required level of continuous support

Consider the staffing levels and skills required to support the CDMS in a sustainable manner.

Consider whether the organization should retain the ability to maintain and enhance the CDMS software and determine the staffing levels and skills required to do so.

What level of investment can the organization support over the long term?

What will the impact be on the long-term viability of the CDMS if these skills are missing?

Can these resources be obtained elsewhere, including from neighbouring countries, development agencies, Regional Climate Centres and so forth (see A Proposed Long Term Resource Plan for Climate Data Management and Rescue Activities,2017, Appendix D).

How can a business case be developed to obtain the required level of continuous support?

10.8 Next steps

10.8.1 Develop business case

Using the information gathered in this chapter, develop an overview of the work required to:

- Implement the CDMS;
- Migrate existing data to the new CDMS;
- Transition the new CDMS to an operational environment;
- Decommission the old CDMS;
- Operate and maintain the CDMS on a sustainable basis over the long term.

Consider also a communication strategy to ensure that all stakeholders (internal and, where relevant, external to the NMHS) are kept aware of developments in the project, and in particular of any developments that may affect service provision.

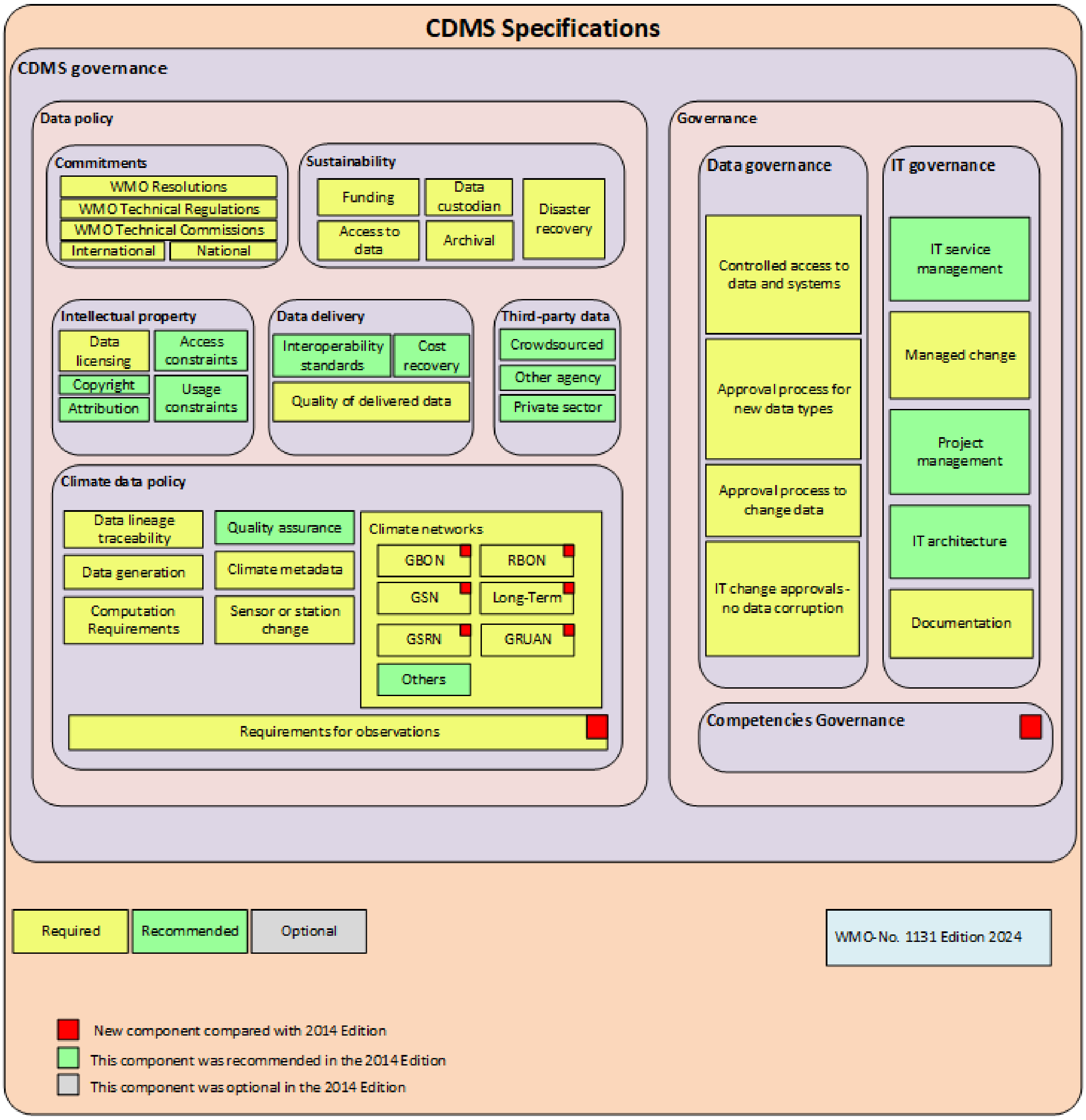
With this information, develop a business case to obtain support from the organization, and if necessary the government or an external agency, to receive appropriate funding to implement and maintain the new CDMS.

10.8.2 Develop and implement project plan

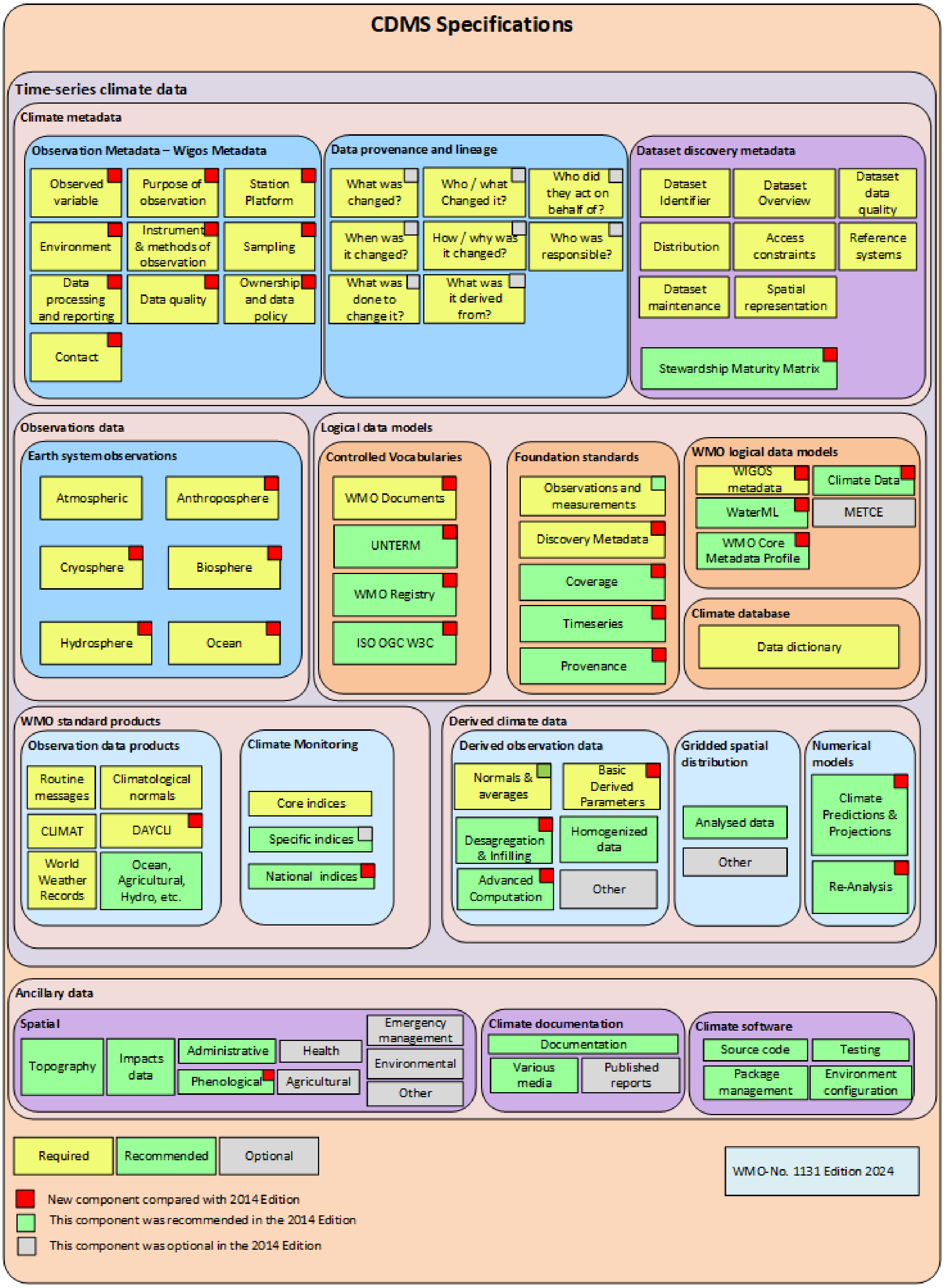
Once organizational support has been obtained and sufficient funding found, implement the new CDMS based on an appropriately resourced IT project.

Appendix 1 Diagrams of climate data management system components

Governance

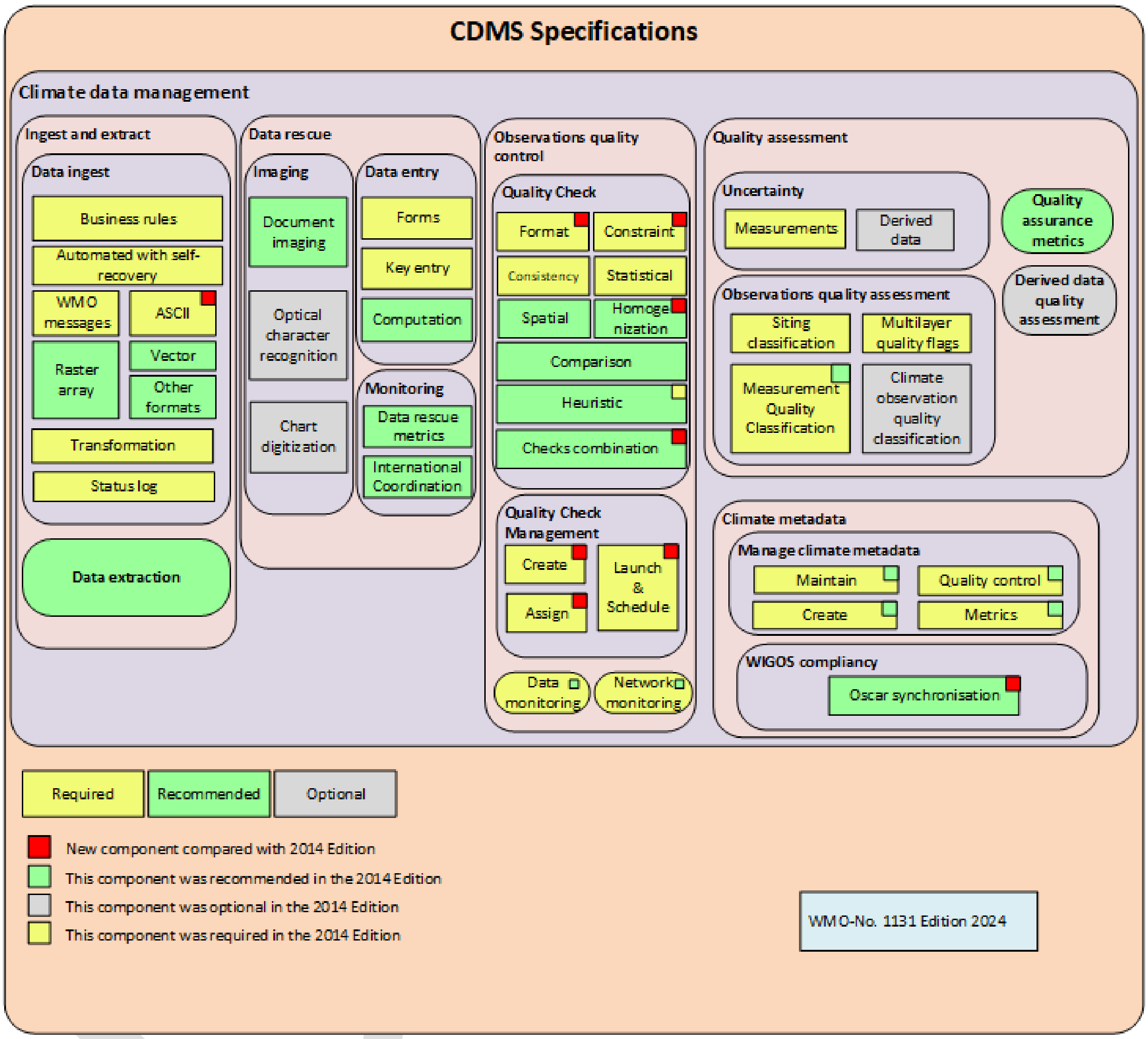


Climate Data

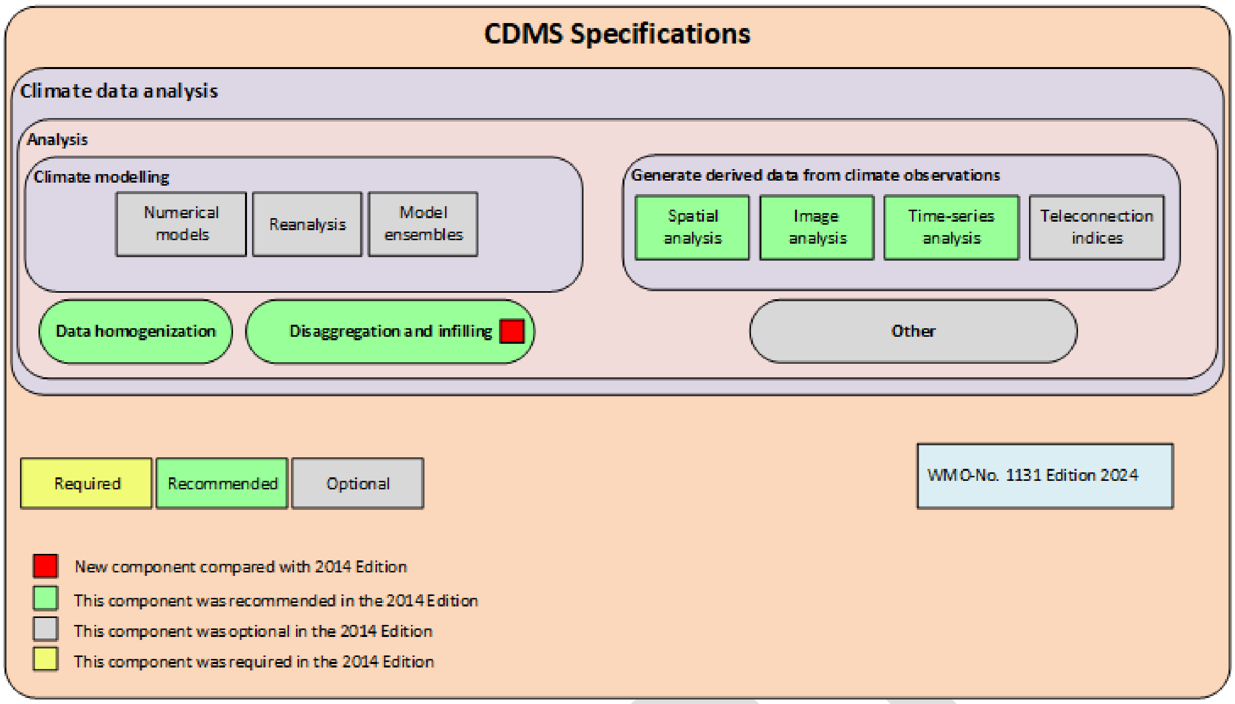

CDMS Specifications
Time-series climate data
Climate metadata
Observation Metadata – Wigos Metadata
Observed variable
Purpose of observation
Station Platform
Environment
Instrument & methods of observation
Sampling
Data processing and reporting
Data quality
Ownership and data policy
Contact
Data provenance and lineage
What was changed?
Who / what Changed it?
Who did they act on behalf of?
When was it changed?
How / why was it changed?
Who was responsible?
What was done to change it?
What was it derived from?
Dataset discovery metadata
Dataset Identifier
Dataset Overview
Dataset data quality
Distribution
Access constraints
Reference systems
Dataset maintenance
Spatial representation
Stewardship Maturity Matrix
Observations data
Earth system observations
Atmospheric
Anthroposphere
Cryosphere
Biosphere
Hydrosphere
Ocean
Logical data models
Controlled Vocabularies
WMO Documents
UNTERM
WMO Registry
ISO OGC W3C
Foundation standards
Observations and measurements
Discovery Metadata
Coverage
Timeseries
Provenance
WMO logical data models
WIGOS metadata
Climate Data
WaterML
METCE
WMO Core Metadata Profile
Climate database
Data dictionary
WMO standard products
Observation data products
Routine messages
Climatological normals
CLIMAT
DAYCLI
World Weather Records
Ocean, Agricultural, Hydro, etc.
Climate Monitoring
Core indices
Specific indices
National indices
Derived climate data
Derived observation data
Normals & averages
Basic Derived Parameters
Desagregation & Infilling
Homogenized data
Advanced Computation
Other
Gridded spatial distribution
Analysed data
Other
Numerical models
Climate Predictions & Projections
Re-Analysis
Ancillary data
Spatial
Topography
Impacts data
Administrative
Health
Emergency management
Environmental
Phenological
Agricultural
Other
Climate documentation
Documentation
Various media
Published reports
Climate software
Source code
Testing
Package management
Environment configuration
Required
Recommended
Optional
WMO-No. 1131 Edition 2024
New component compared with 2014 Edition
This component was recommended in the 2014 Edition
This component was optional in the 2014 Edition

1

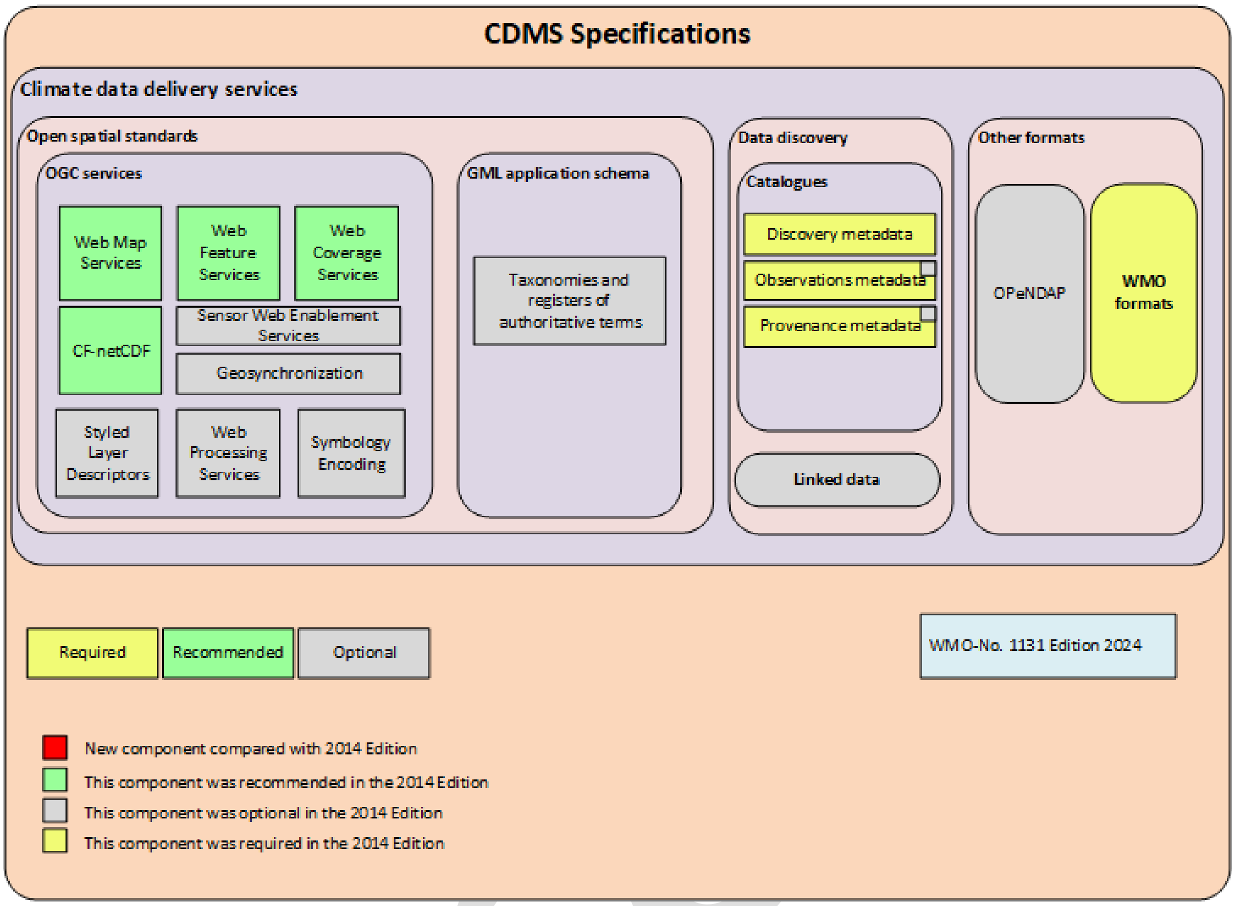
Climate Data Management



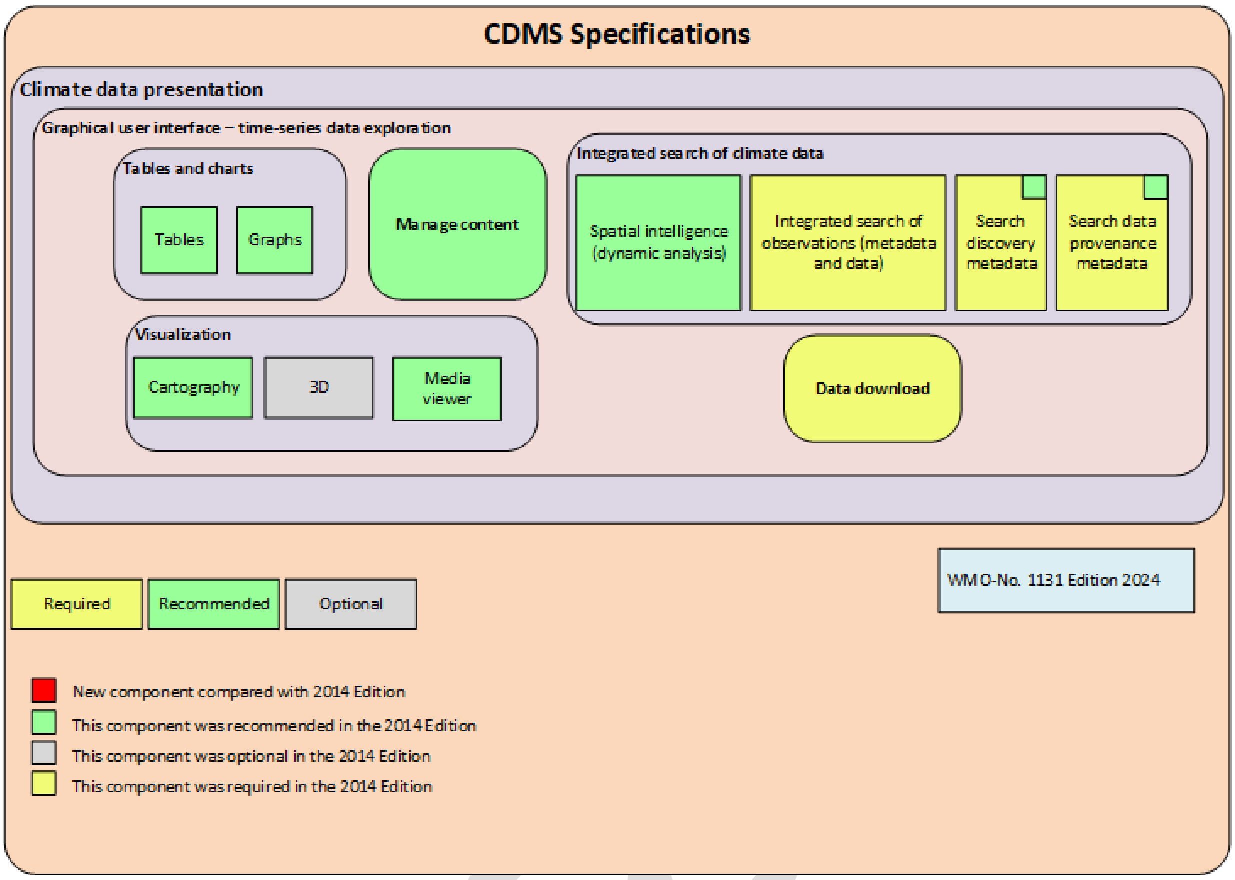
Climate Data Analysis



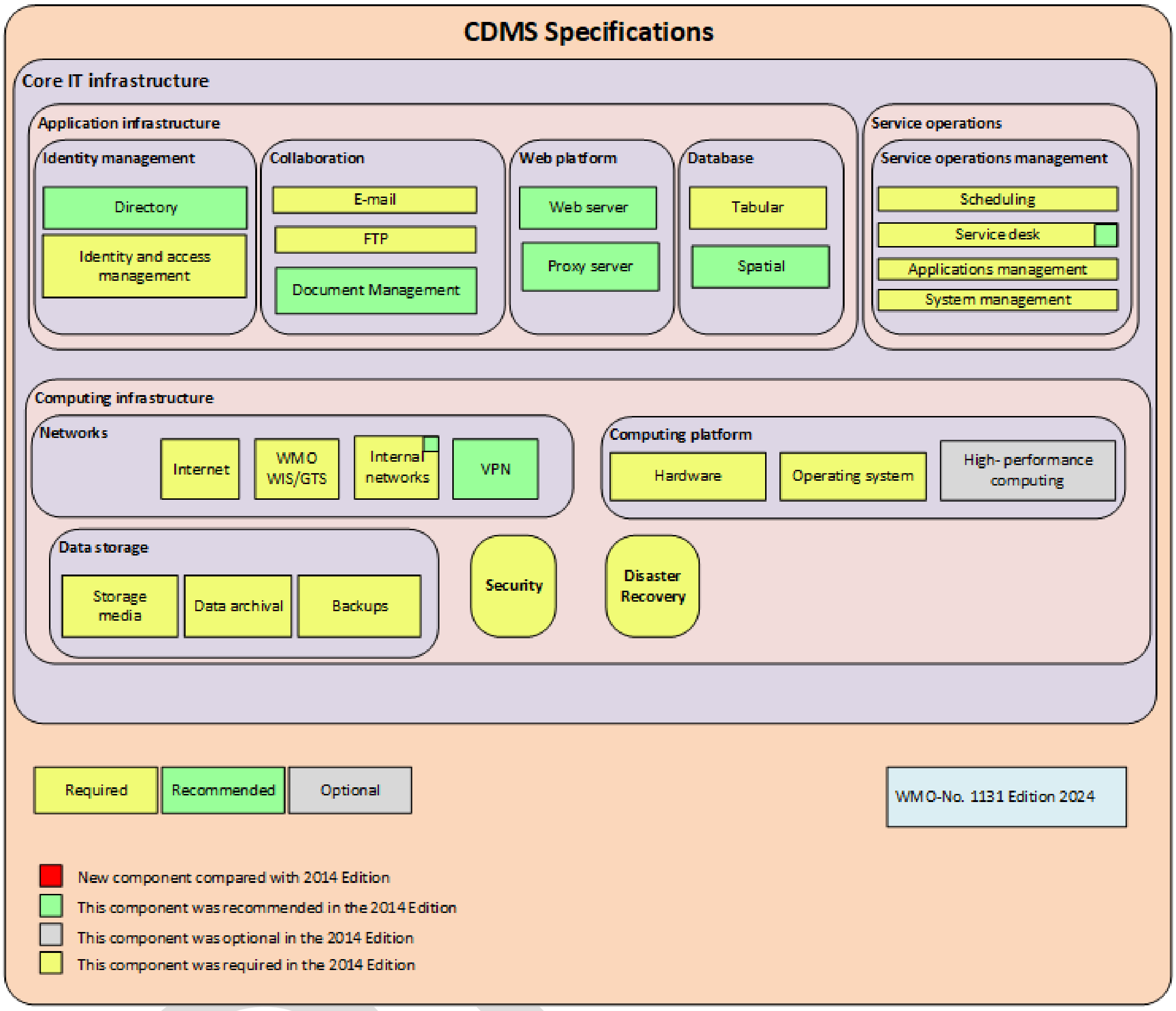
Climate Data Delivery



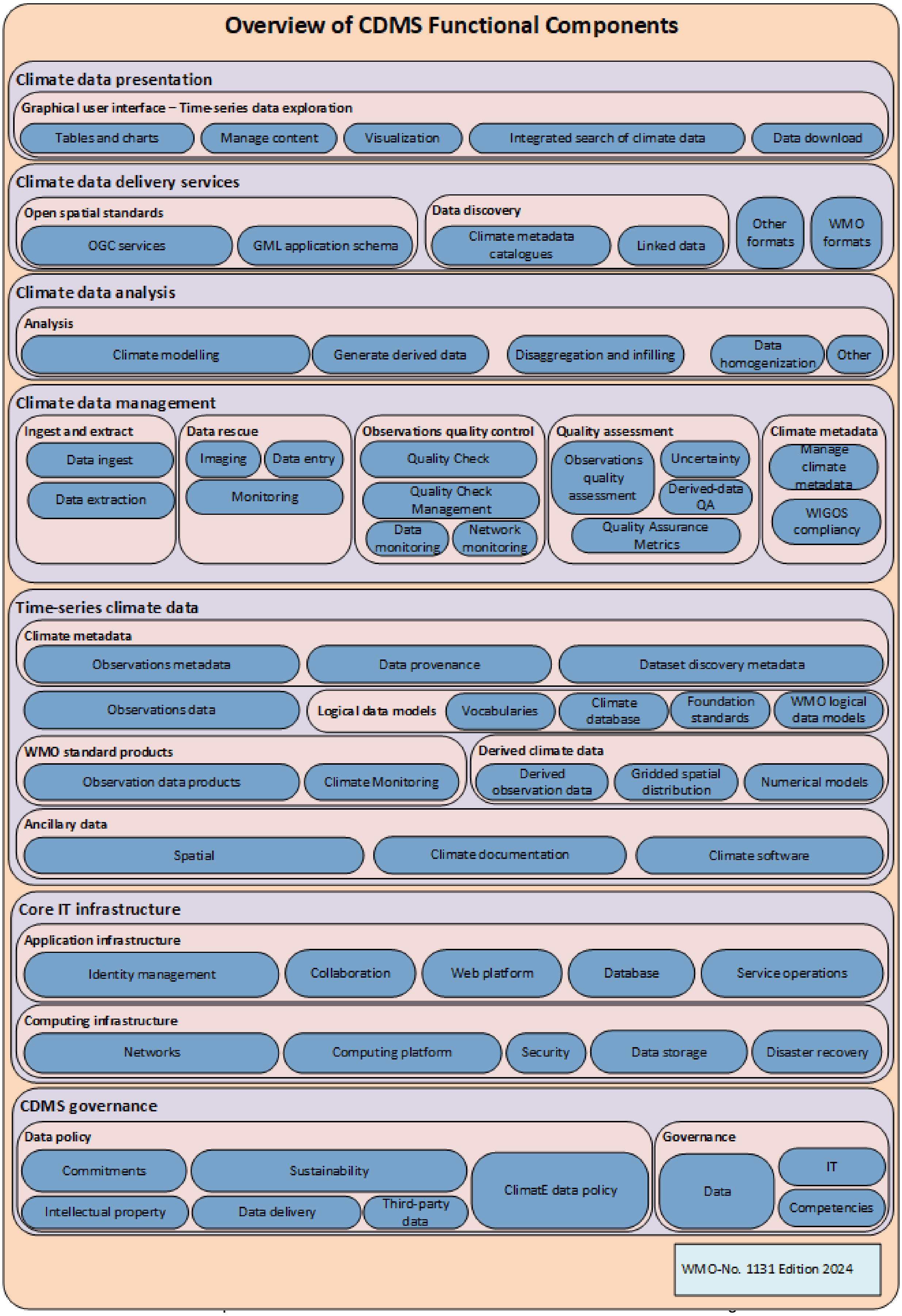
Climate Data Presentation



Climate Data Information Technology



Climate Data Main Components


Overview of CDMS Functional Components
Climate data presentation
Graphical user interface – Time-series data exploration
Tables and charts
Manage content
Visualization
Integrated search of climate data
Data download
Climate data delivery services
Open spatial standards
OGC services
GML application schema
Data discovery
Climate metadata catalogues
Linked data
Other formats
WMO formats
Climate data analysis
Analysis
Climate modelling
Generate derived data
Disaggregation and infilling
Data homogenization
Other
Climate data management
Ingest and extract
Data ingest
Data extraction
Data rescue
Imaging
Data entry
Monitoring
Observations quality control
Quality Check
Quality Check Management
Data monitoring
Network monitoring
Quality assessment
Observations quality assessment
Uncertainty
Derived-data QA
Quality Assurance Metrics
Climate metadata
Manage climate metadata
WIGOS compliancy
Time-series climate data
Climate metadata
Observations metadata
Data provenance
Dataset discovery metadata
Observations data
Logical data models
Vocabularies
Climate database
Foundation standards
WMO logical data models
WMO standard products
Observation data products
Climate Monitoring
Derived climate data
Derived observation data
Gridded spatial distribution
Numerical models
Ancillary data
Spatial
Climate documentation
Climate software
Core IT infrastructure
Application infrastructure
Identity management
Collaboration
Web platform
Database
Service operations
Computing infrastructure
Networks
Computing platform
Security
Data storage
Disaster recovery
CDMS governance
Data policy
Commitments
Sustainability
ClimatE data policy
Intellectual property
Data delivery
Third-party data
Governance
Data
IT
Competencies
WMO-No. 1131 Edition 2024

Appendix 2 Summary on components classification

3	**CDMS Governance**	
3.1	Data policy	
3.1.1	Commitments	
3.1.1.1 WMO Resolutions		Required
3.1.1.2 WMO Technical Regulations		Required
3.1.1.3 WMO Technical Commissions		Required
3.1.1.4 International		Required
3.1.1.5 National		Required
3.1.2	Sustainability	
3.1.2.1 Disaster recovery		Required
3.1.2.2 Funding		Required
3.1.2.3 Data custodian		Required
3.1.2.4 Access to data		Required
3.1.2.5 Archival policy		Required
3.1.3	Intellectual property	
3.1.3.1 Data licensing		Required
3.1.3.2 Access constraints		Recommended
3.1.3.3 Usage constraints		Recommended
3.1.3.4 Copyright		Recommended
3.1.3.5 Attribution		Recommended
3.1.4	Data delivery	
3.1.4.1 Interoperability standards		Recommended
3.1.4.2 Quality of delivered data		Required
3.1.4.3 Cost recovery		Recommended
3.1.5	Third-party data	
3.1.5.1 Crowdsourced		Recommended
3.1.5.2 Other agency		Recommended
3.1.5.3 Private sector		Recommended
3.1.6	Climate data policy	
3.1.6.1 Climate metadata		Required
3.1.6.2 Observations Metadata		Required
3.1.6.3 Data lineage traceability		Required

3.1.6.4 Data generation		Required
3.1.6.5 Climate networks		Required
3.1.6.6 Requirements for observations		Required
3.1.6.7 Sensor or station change		Required
3.1.6.8 Quality assurance		Recommended
3.2	Governance	
3.2.1	Data governance	
3.2.1.1 Controlled access to data and systems		Required
3.2.1.2 Approval process for new data types		Required
3.2.1.3 Approval process to change data		Required
3.2.1.4 IT change approvals – no data corruption		Required
3.2.2	IT governance	
3.2.2.1 IT service management		Recommended
3.2.2.2 Managed change		Required
3.2.2.3 Project management		Recommended
3.2.2.4 IT architecture		Recommended
3.2.2.5 Documentation		Required
3.2.3	Competencies governance	
3.2.3.1 Competencies governance		Optional
4	**Time-series climate data**	
4.1	Observations data	
4.1.1.1 Atmosphere		Required
4.1.1.2 Ocean		Required
4.1.1.3 Cryosphere		Required
4.1.1.4 Antrophosphere		Required
4.1.1.5 Biosphere		Required
4.1.1.6 Hydrosphere		Required
4.2	Logical data models	
4.2.1	Controlled Vocabularies	
4.2.1.1 WMO Documents		Required
4.2.1.2 WMO Codes Registry		Recommended
4.2.1.3 UNTERM		Recommended
4.2.1.4 ISO OGC W3C		Recommended
4.2.2	Foundation standards	
4.2.2.1 Observations, Measurements and Samples		Required
4.2.2.2 Coverage		Recommended
4.2.2.3 Timeseries		Recommended
4.2.2.4 Provenance		Recommended
4.2.2.5 Discovery Metadata		Required
4.2.3	WMO logical data models	

4.2.3.1 WIGOS Metadata		Required
4.2.3.2 Climate Data		Recommended
4.2.3.3 WMO Core Metadata Profile (WCMP)		Required
4.2.3.4 WaterML		Recommended
4.2.3.5 METCE		Optional
4.2.4	Climate database	
4.2.4.1 Data dictionary		Required
4.3	Climate metadata	
4.3.1	Observations metadata	
4.3.1.1 Observed variable		Required
4.3.1.2 Purpose of observation		Required
4.3.1.3 Station/platform		Required
4.3.1.4 Environment		Required
4.3.1.5 Instruments and methods of observation		Required
4.3.1.6 Sampling		Required
4.3.1.7 Data processing and reporting		Required
4.3.1.8 Data quality		Required
4.3.1.9 Ownership and data policy		Required
4.3.1.10 Contact		Required
4.3.2	Dataset discovery metadata	
4.3.2.1 Dataset identifier		Required
4.3.2.2 Dataset overview		Required
4.3.2.3 Dataset data quality		Required
4.3.2.4 Distribution		Required
4.3.2.5 Access constraints		Required
4.3.2.6 Dataset maintenance		Required
4.3.2.7 Spatial representation		Required
4.3.2.8 Reference systems		Required
4.3.2.9 Stewardship Maturity Matrix		Recommended
4.3.3	Data provenance and Lineage	
4.3.3.1 What was changed?		Required
4.3.3.2 When was it changed?		Required
4.3.3.3 What was it derived from?		Required
4.3.3.4 What was done to change it?		Required
4.3.3.5 How/why was it changed?		Required
4.3.3.6 Who/what changed it?		Required
4.3.3.7 Who did they act on behalf of?		Required
4.3.3.8 Who was responsible?		Required
4.4	WMO standard products	
4.4.1	Observation data products	
4.4.1.1 Routine messages		Required

4.4.1.2 Climatological normals		Required
4.4.1.3 CLIMAT		Required
4.4.1.4 DAYCLI		Required
4.4.1.5 World Weather Records		Required
4.4.1.6 Ocean, Agricultural, Hydro, Aeronautical etc.		Recommended
4.4.2	Climate Monitoring	
4.4.2.1 Core indices		Required
4.4.2.2 Specific indices		Recommended
4.4.2.3 National indices and monitoring products		Recommended
4.5	Derived climate data	
4.5.1	Derived observation data	
4.5.1.1 Normals and averages		Required
4.5.1.2 Basic derived variables		Required
4.5.1.3 Homogenized data		Recommended
4.5.1.4 Advanced Calculation		Recommended
4.5.1.5 Other		Optional
4.5.2	Gridded spatial distribution of observations	
4.5.2.1 Analysed data		Recommended
4.5.2.2 Other		Optional
4.5.3	Numerical models	
4.5.3.1 Climate Predictions and projections		Recommended
4.5.3.2 Reanalysis		Recommended
4.6	Ancillary data	
4.6.1	Spatial	
4.6.1.1 Topography		Recommended
4.6.1.2 Emergency management		Optional
4.6.1.3 Agricultural		Optional
4.6.1.4 Health		Optional
4.6.1.5 Environmental		Optional
4.6.1.6 Phenological data		Recommended
4.6.1.7 Administrative		Recommended
4.6.1.8 Impacts data		Recommended
4.6.1.9 Other		Optional
4.6.2	Climate documentation	
4.6.2.1 Published reports		Optional
4.6.2.2 Documentation		Recommended
4.6.2.3 Various media		Recommended
4.6.3	Climate software	
4.6.3.1 Source code management		
4.6.3.2 Package management		

4.6.3.3 Environment configuration		
4.6.3.4 Software testing		
5	**Climate data management**	
5.1	Ingest and extract	
5.1.1	Data ingest	
5.1.1.1 Business rules		Required
5.1.1.2 WMO messages		Required
5.1.1.3 Vector		Recommended
5.1.1.4 Raster array		Recommended
5.1.1.5 ASCII		Required
5.1.1.6 Other formats		Recommended
5.1.1.7 Status log		Required
5.1.1.8 Automated with self-recovery		Required
5.1.1.9 Transformation		Required
5.1.2	Data extraction	
5.1.2.1 Data extraction		Recommended
5.2	Data rescue	
5.2.1	Imaging- & digitization	
5.2.1.1 Document imaging		Recommended
5.2.1.2 Optical character recognition		Optional
5.2.1.3 Chart digitization		Optional
5.2.2	Monitoring	
5.2.2.1 Data rescue metrics		Recommended
5.2.2.2 International Coordination		Recommended
5.2.3	Data entry, also known as key entry	
5.2.3.1 Forms		Required
5.2.3.2 Key entry		Required
5.2.3.3 Computation		Recommended
5.3	Observations quality control	
5.3.1	Quality checking	
5.3.1.1 Format checks		Required
5.3.1.2 Constraint checks		Required
5.3.1.3 Consistency checks		Required
5.3.1.4 Statistical checks		Required
5.3.1.5 Spatial checks		Recommended
5.3.1.6 Homogenization checks		Optional
5.3.1.7 Comparison checks		Recommended
5.3.1.8 Heuristic checks		Recommended
5.3.1.9 Checks combination		Recommended
5.3.2	Quality Check Management	
5.3.2.1 Create and Modify		Required
5.3.2.2 Assign variables		Required
5.3.2.3 Launching and Scheduling		Required

5.3.2.4 Network monitoring		Recommended
5.3.3	Data monitoring	
5.3.3.1 Data monitoring		Required
5.4	Quality assessment	
5.4.1	Observations quality assessment	
5.4.1.1 Siting classification		Required
5.4.1.2 Measurement Quality Classification		Required
5.4.1.3 Multilayer quality flags		Required
5.4.1.4 Climate observation quality classification		Optional
5.4.2	Derived-data quality assessment	
5.4.2.1 Derived-data quality assessment		Optional
5.4.3	Quality assurance metrics	
5.4.3.1 Quality assurance metrics		Recommended
5.4.4	Uncertainty	
5.4.4.1 Measurements		Required
5.4.4.2 Derived data		Recommended
5.5	Climate metadata	
5.5.1	Manage climate metadata	
5.5.1.1 Create		Required
5.5.1.2 Maintain		Required
5.5.1.3 Quality control		Required
5.5.1.4 Metrics		Required
5.5.2	WIGOS compliancy	
5.5.2.1 OSCAR Synchronisation		Recommended
5.6	References	
6	**Climate Data Analysis**	
6.1	Analysis	
6.1.1	Climate Modelling	
6.1.1.1 Numerical Models		Optional
6.1.1.2 Reanalysis		Optional
6.1.1.3 Model Ensembles		Optional
6.1.2	Generate Derived Data from Climate Observations	
6.1.2.1 Spatial Analysis		Recommended
6.1.2.2 Image Analysis		Recommended
6.1.2.3 Time Series Analysis		Recommended
6.1.2.4 Teleconnection Indices		Optional
6.1.3	Data Homogenisation	
6.1.3.1 Data Homogenisation		Recommended
6.1.4	Disaggregation and infilling	
6.1.4.1 Disaggregation and infilling		Recommended

6.1.5	Other	
6.1.5.1 Other		
7	**Climate data presentation**	
7.1	Graphical user interface – time-series data exploration	
7.1.1	Tables and charts	
7.1.1.1 Tables		Recommended
7.1.1.2 Graphs		Recommended
7.1.2	Manage content	
7.1.2.1 Manage content		Recommended
7.1.3	Visualization	
7.1.3.1 Cartography		Recommended
7.1.3.2 3D		Optional
7.1.3.3 Media viewer		Recommended
7.1.4	Integrated search of climate data	
7.1.4.1 Spatial intelligence		Recommended
7.1.4.2 Integrated search of observations		Required
7.1.4.3 Search discovery metadata		Required
7.1.4.4 Search data provenance metadata		Required
7.1.5	Data download	
7.1.5.1 Data download		Recommended
8	**Climate data delivery services**	
8.1	Open spatial standards	
8.1.1	Open Geospatial Consortium services	
8.1.1.1 Web Map Services		Recommended
8.1.1.2 Web Feature Services		Recommended
8.1.1.3 Web Coverage Services		Recommended
8.1.1.4 Sensor Web Enablement Services		Optional
8.1.1.5 CF-netCDF		Recommended
8.1.1.6 Geosynchronization		Optional
8.1.1.7 Web Processing Services		Optional
8.1.1.8 Symbology Encoding		Optional
8.1.1.9 Styled Layer Descriptors		Optional
8.1.2	Geography Markup Language application schema	
8.1.2.1 Taxonomies & registers of authoritative terms		Optional
8.2	Data discovery	
8.2.1	Climate metadata catalogues	
8.2.1.1 Discovery metadata catalogue		Required
8.2.1.2 Observations metadata catalogue		Required
8.2.1.3 Data provenance metadata catalogue		Required

8.2.2	Linked data	
8.2.2.1 Linked data		Optional
8.3	Other formats	
8.3.1 WMO Formats		Required
9	**Core IT infrastructure**	
9.1	Application infrastructure	
9.1.1	Identity management	
9.1.1.1 Directory		Recommended
9.1.1.2 Identity and access management		Required
9.1.2	Collaboration	
9.1.2.1 E-mail		Required
9.1.2.2 FTP		Required
9.1.2.3 Document management		Recommended
9.1.3	Web platform	
9.1.3.1 Web server		Recommended
9.1.3.2 Proxy server		Recommended
9.1.4	Database	
9.1.4.1 Tabular		Required
9.1.4.2 Spatial		Recommended
9.2	Service operations	
9.2.1	Service operations management	
9.2.1.1 Scheduling		Required
9.2.1.2 Service desk		Required
9.2.1.3 Applications management		Required
9.2.1.4 Systems management		Required
9.3	Computing infrastructure	
9.3.1	Networks	
9.3.1.1 Internet		Required
9.3.1.2 WMO WIS/GTS		Required
9.3.1.3 Internal networks		Required
9.3.1.4 VPN		Recommended
9.3.2	Computing platform	
9.3.2.1 Hardware		Required
9.3.2.2 Operating system		Required
9.3.2.3 High-performance computing		Recommended
9.3.3	Security	
9.3.3.1 Security		Required
9.3.4	Data storage	
9.3.4.1 Storage media		Required
9.3.4.2 Data archival		Required
9.3.4.3 Backups		Required
9.3.5	Disaster recovery	
9.3.5.1 Disaster recovery		Required