WORLD METEOROLOGICAL ORGANIZATION
COMMISSION FOR CLIMATOLOGY

*__IMPORTANT NOTE__ as of 10 June 2014: **This draft version has been made available for preparation of CCI-XVI, which is expected to endorse the CDMS Specifications publication (cf. Doc. 8.1, CCI-XVI). This draft version of CCI's ET-CDMS has been reviewed by international experts and underwent a preliminary editing by WMO. Final editing and layout is underway. This draft version will be replaced by the final WMO publication once available.***

Climate Data Management System Specifications
Version 1.2

Contents

Acknowledgements

1   Executive summary

This publication establishes a framework defining the functionality required within a climate data management system (CDMS). A CDMS is an integrated computer-based system that facilitates the effective 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.

Not all components are needed within every CDMS. Components are classified as required (i.e. mandatory), recommended (i.e. 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.

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.

It is expected that the World Meteorological Organization (WMO) and NMHSs will take some time to make the necessary changes to their CDMS. Therefore, a period of three to five years should be allowed for organizations to meet the required functionality. Chapter 10 outlines an approach that organizations could take to make this change.

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.

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 CDMSs.

This publication defines a set of policies and governance processes that are necessary to effectively manage climate data. These policies should be implemented as a global framework to facilitate better integration of climate data between NMHSs and ease the workload required for regional and global analysis of climate data.

To this end, the publication recommends the establishment of a global Climate Data Framework. Similar proposals are currently being discussed, such as the High Quality Global Data Management Framework for Climate. These concepts will be coordinated to ensure that only a single authoritative framework is implemented.

It is expected that this publication will become a live document, with updated versions released as required. Recommendation 11.2 outlines some of the future work required.

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

Definitions

The climate archive	The climate archive is a set of data which is useful in the provision of value-added climate products, analyses or services. The archive consists of all the data and information defined in Chapter 4 of this publication and includes the climate record (see definition below). The climate archive should be managed as a permanent repository of climate-relevant data.
The climate record	The climate record is the "official" climate data repository for an NMHS. It contains 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 • Homogenized data • Satellite imagery • Radar data The data comprising the record consist of observation values, along with values following routine quality-control processes. The climate record should be managed as a permanent repository of climate-relevant data.
Authoritative data source	An authoritative data source is the accepted source of truth for the data.

2 Background

This publication has been produced as a result of numerous requests for guidance on which CDMS should be implemented by organizations to effectively manage and utilize climate data.

While many systems claim to be a CDMS, the functionality expected within such CDMSs has yet to be clearly defined. Traditionally, CDMSs have been regarded as a black box, with the internal workings and components hidden from view. This publication is an attempt to reverse this situation and provide the first formal definition of the functionality expected within a CDMS.

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 underpin future work to compare the functionality available in competing CDMSs.

It is anticipated that this Specification will become a living document, to which more detailed definitions of requirements will be added as they become available. Recommendation 11.2 outlines additional work that is required on this Specification.

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.

There are very few currently available systems, if any, that will meet the requirements as defined in this Specification. This is not considered a problem as CDMSs have traditionally been treated as a black-box software application, with no consistency as to CDMS requirements. This will now change with this Specification.

Climate and CDMS procurement staff of NMHSs, climate data managers, systems integrators and CDMS developers, architects, vendors and so forth are advised to:

- Treat this Specification as a description of the functionality that is either required or desirable for their CDMS.

- Note that a CDMS is not expected to contain all of the functionalities described in this Specification.

- Note that a CDMS is not expected to contain all of its functionality within 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. Some effort, however, will be required to integrate these components together.

- 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 below

    – 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."[1]

---

[1] 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 1 is a simplified pictorial representation of many of the aspects that make up a climate system.
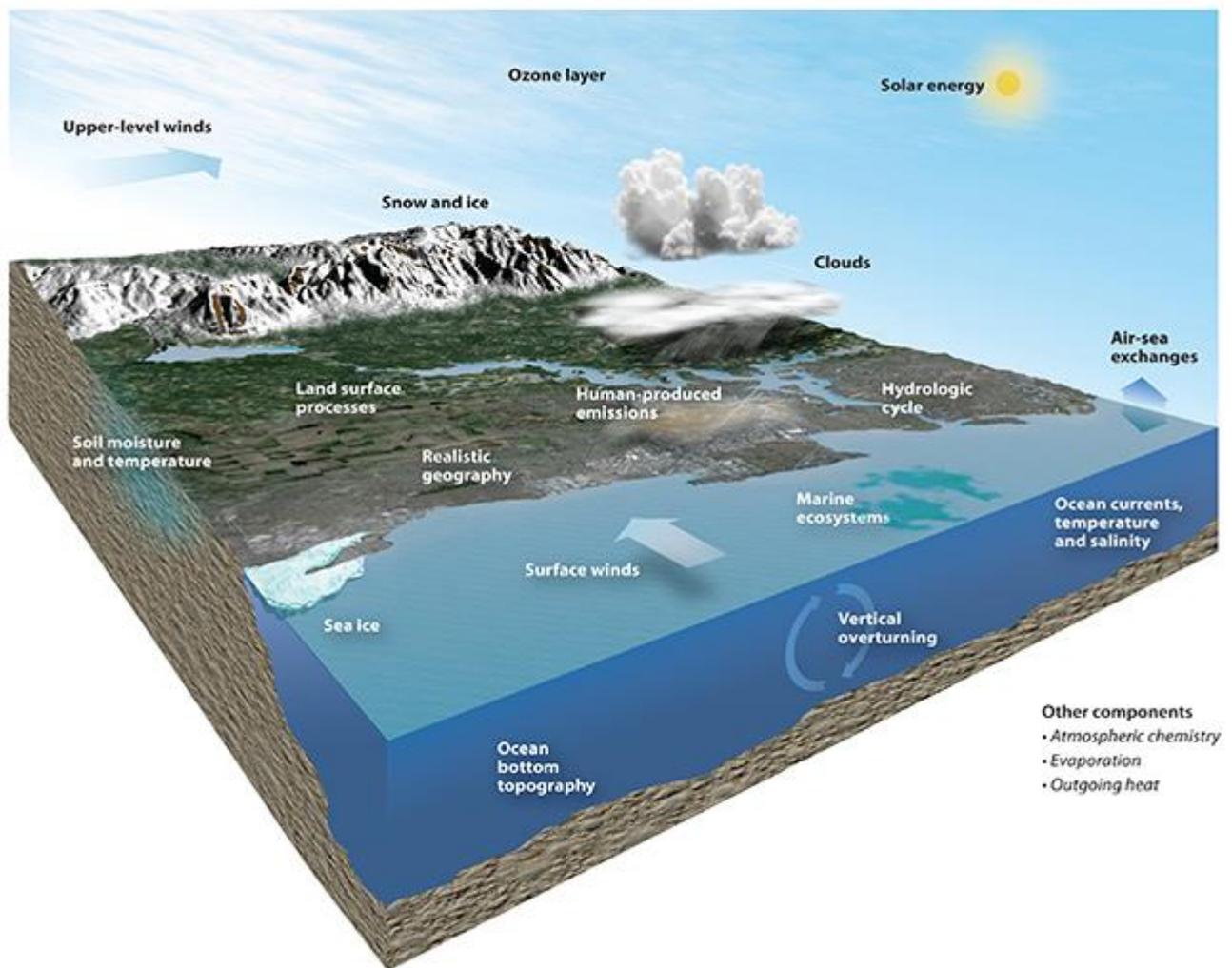
Figure 1.Community Earth System Model

Source: National Center for Atmospheric Research, United States of America. Image can be found at http://www2.ucar.edu/sites/default/files/news/2011/CESM_final.jpg.

For a more detailed overview of the climate system, see:
- *Guide to Climatological Practices* (WMO-No. 100)
- Wikipedia article on climate

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.

Figure 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 models. 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.

For a more detailed overview, see:
- *Guide to Climatological Practices* (WMO-No. 100)
- *Guidelines on Climate Observation Networks and Systems* (WMO/TD-No. 1185), WCDMP-52
- WMO Integrated Global Observing System (WIGOS) flyer

2.3 Major components of a climate data management system

Figure 3 is a graphic depiction of the major functional components that comprise a CDMS.
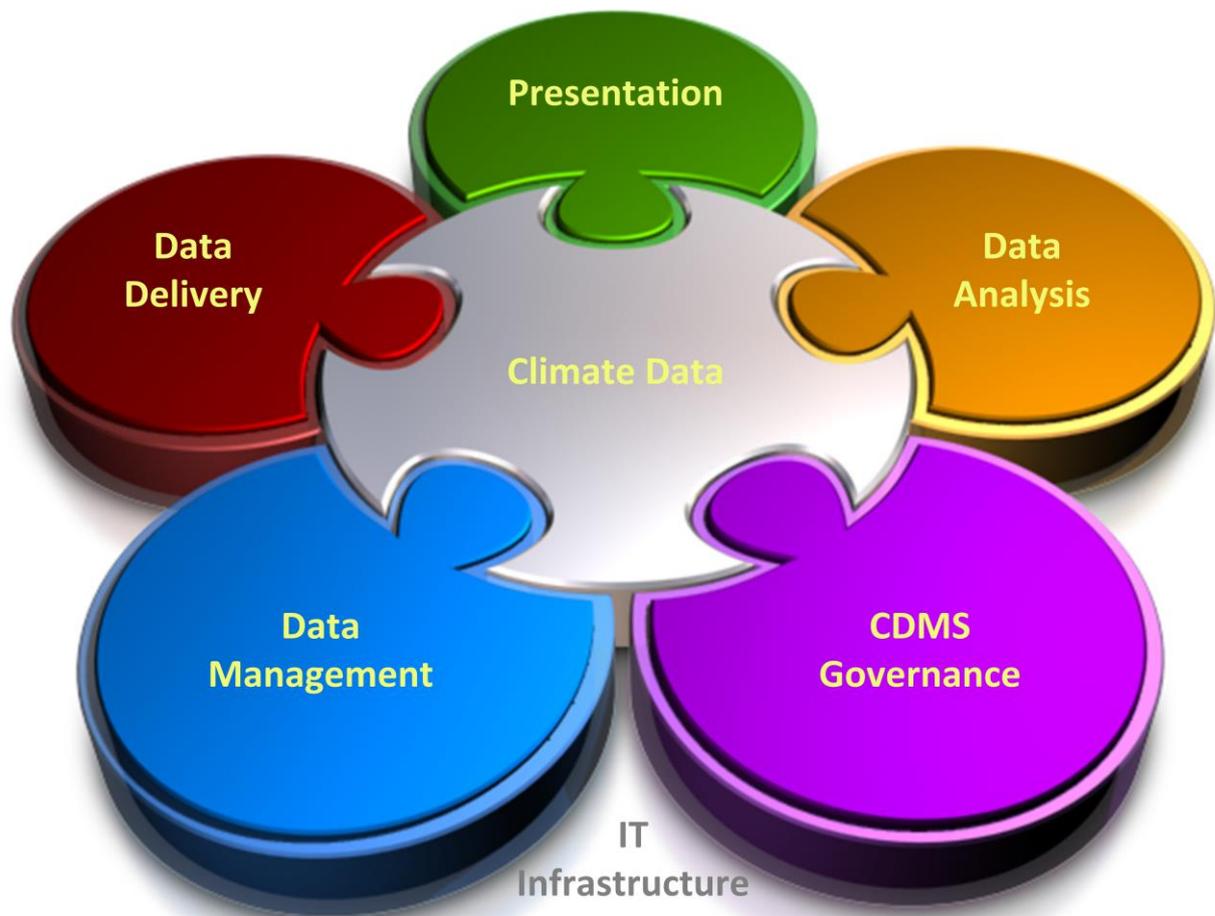
High Level Components of a CDMS

Figure 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:

- Global Climate Observing System (GCOS) Essential Climate Variables (ECVs)

- An expanded view of climate metadata that includes metadata on observations, discovery and data provenance
- 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 data, documentation and climate software
- Other important data such as logical data models

2.3.2 Governance of climate data management systems

The CDMS governance component refers to a consistent set of policies and governance processes needed to build a solid foundation for the establishment and management of authoritative sources of climate data and related services. Although a number of WMO initiatives will establish consistent policies in due course, this component may help to immediately improve many NMHS data management practices.

This component contains the following concepts:
- Data policy, including:
  - Organizational commitments
  - Ensuring the sustainability of CDMSs
  - Intellectual property
  - Data delivery
  - Third-party data
  - Climatology policy
- Governance, including:
  - Data governance
  - IT 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
- Quality assessment
- Management of climate metadata

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:

- A series of techniques, including statistical, spatial and image analysis
- Homogenization
- Numerical modelling processes

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 IT infrastructure

The IT 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 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.

Overview of CDMS Functional Components

Climate Data Presentation

Printed Media
- Written Reports

Graphical User Interface – Time Series Data Exploration
- Business Intelligence Reports
- Manage Content
- Visualisation
- Integrated Search of Climate Data
- Data Download

Climate Data Delivery Services

Open Spatial Standards
- OGC Services
- GML Application Schema

Data Discovery
- Climate Metadata Registers

- Other Formats
- WMO Formats

Climate Data Analysis

Analysis
- Climate Modelling
- Generate Derived Data from Climate Observations
- Data Homogenisation
- Other

Climate Data Management

Extract Transform Load
- Data Extraction
- Data Ingest

Data Rescue
- Imaging
- Data Entry
- Monitoring

Observations Quality Control
- Quality Management
- Data Monitoring
- Network Monitoring

Quality Assessment
- Observations Quality Assessment
- Uncertainty
- Derived Data QA

Climate Metadata
- Manage Climate Metadata

Time Series Climate Data

Climate Metadata
- Observations Metadata
- Data Provenance
- Dataset Discovery Metadata

- Observations Data
- **Logical Data Models**
- Climate Database
- Foundation Standards
- WMO Logical Data Models

WMO Standard Products
- Observation Data Products
- Climate Change Indices

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 Management

Computing Infrastructure
- Networks
- Computing Platform
- Security
- Data Storage
- Disaster Recover

CDMS Governance

Data Policy
- Commitments
- Sustainability
- Intellectual Property
- Data Delivery
- Third Party Data
- Climatology Policy

Governance
- Data Governance
- IT Governance

- Training
- System enhancement capability

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

Figure 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. A single component may refer to a range of software and processes that provide a functional requirement. 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 as required, recommended or optional. This classification scheme and associated definitions may be found in section 2.6 of this publication.

NMHSs can use the CDMS component classification scheme to help them determine what functionality their organization can afford to support over the long term.

In addition, NMHSs may find that more advanced CDMS functionalities may be available via other sources, such as Regional Climate Centres and the WMO Information System (WIS).

2.5.1.2 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 that is expected to be required of future CDMSs within a five-year time frame. This can be thought of as the future state of the CDMS.

It is anticipated that NMHSs will 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 afford to sustain in the long term to implement, support and maintain their CDMS functionality. This investment will not be restricted to just IT 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 IT skills.

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

Taking into account the amount 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 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 this classification may change with future revisions of this publication.

Required	Recommended	Optional

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).	If one of the required functions is not fulfilled, the system is not complying with WMO requirements and should therefore not be considered as a WMO-compliant CDMS.
Recommended	This component's functionality is recommended in order to comply with current best-practices.	Due to the evolving nature of the CDMS, these 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.

3 Governance of climate data management systems
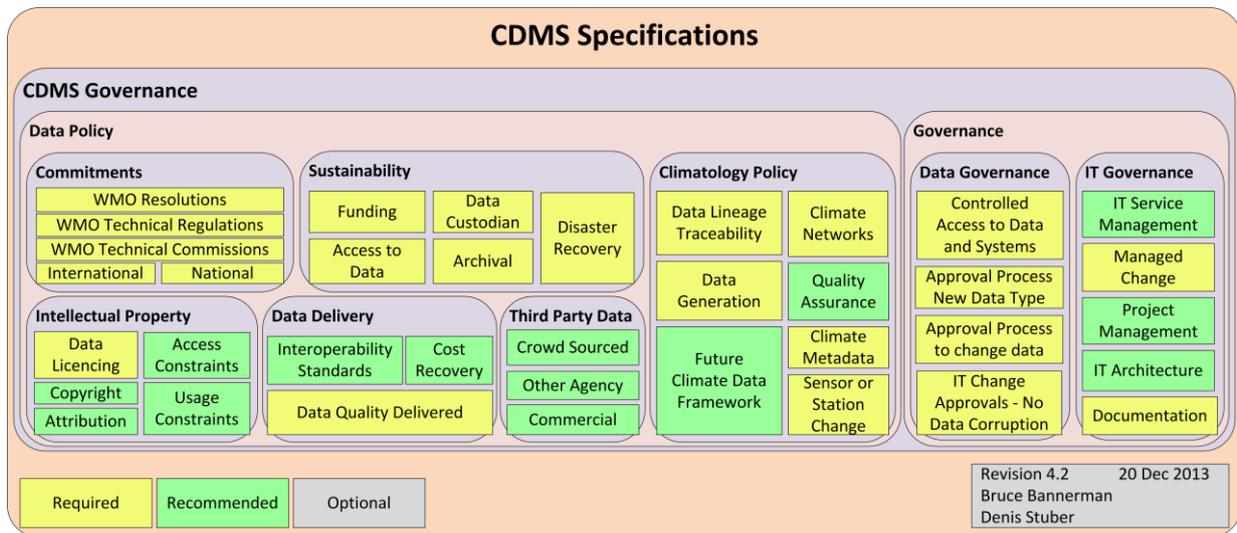
Figure 5. Governance of climate data management systems

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 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.

Even with a robust data policy framework in place, issues will still arise that will need to be handled on a case-by-case basis, though these cases will be minimized.

CDMS data policy and governance is an area that requires considerable development to ensure consistent approaches across organizations 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 enhance data integrity.
- Ensure the long-term sustainability of NMHS climate records.

In the recommendations section, readers will find a related suggestion to establish the WMO Climate Data Framework that provides, in part, consistent policies, requirements and standards for NMHSs to adopt.

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

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). Some resolutions have major implications for data policy, such as Resolution 40 (Cg-XII), which provides guidelines for the policy and practice of exchanging meteorological data, and Resolution 25 (Cg-XIII), which provides guidelines on the exchange of hydrological data. The major innovation in these resolutions is the definition of essential data. Through these resolutions, WMO Members commit to making essential data available on a free and unrestricted basis at no more than the reasonable cost of extraction and formatting for delivery to the user. Members may not impose additional charges for essential data.	**Required**
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 four 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 • Volume IV – Quality Management Also see the publications of the annexes that are part of the Technical Regulations: • Annex I – *International Cloud Atlas* (WMO-No. 407), Volume I	**Required**

	- Annex II – *Manual on Codes* (WMO-No. 306), Volumes I.1 and I.2 - Annex III – *Manual on the Global Telecommunication System* (WMO-No. 386), Volume I - Annex IV – *Manual on the Global Data-processing and Forecasting System* (WMO-No. 485), Volume I - Annex V – *Manual on the Global Observing System* (WMO-No. 544), Volume I - Annex VI – *Manual on Marine Meteorological Services* (WMO-No. 558), Volume I - Annex VII – *Manual on the WMO Information System* (WMO-No. 1060) - Annex VIII – *Manual on the Implementation of Education and Training Standards in Meteorology and Hydrology* (WMO-No. 1083), Volume I	
3.1.1.3 **WMO technical commission guides**	This component includes the recommendations or best practices that are generally drawn up as guides published by the different technical commissions: - Commission for Basic Systems - Commission for Instruments and Methods of Observation - Commission for Hydrology - Commission for Atmospheric Sciences - Commission for Aeronautical Meteorology - Commission for Agricultural Meteorology - Commission for Climatology - Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM) While most of the guides are relevant to CDMSs, it is recommended that particular attention be paid to: *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), from the Commission for Instruments and Methods of Observation *Guide to Climatological Practices* (WMO-No. 100), from the Commission for Climatology	**Required**
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,	**Required**

	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 open spatial standards. Note that not all standards are relevant.	
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. • National spatial data infrastructure frameworks that govern the exchange of national spatial data.	<mark>Required</mark>

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?
	    – In the interim, while NMHSs are
		<mark>Required</mark>

	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 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?	

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.	**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.	**Required**

	Some examples of a data custodian's duties are: • Preserving the integrity of the climate record, including quality control and ensuring that observational networks 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 the 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? 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	**Required**	

	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 perhaps the requirement to acknowledge the data source and ensure that the data are made available under a share-alike 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. 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

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 understood. For example: • What data licences apply to NMHS data? • What data licences will the NMHS release its data under? For example: – 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?	Required
3.1.3.2 **Access**	This component concerns policies that clearly define any access constraints relating to the use of climate data.	Recommended

constraints	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?	
3.1.3.3 **Usage constraints**	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. • 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.	Recommended
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

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 available 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. In reality though, only NMHSs and closely related organizations will be able to understand the formats and have the software to use the data. For more information, see section 8.1 of this publication.	Recommended
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 quality of climate data does the organization undertake to deliver and in what circumstances? • Will the organization only provide high-quality 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 25 (Cg-XIII) and 40 (Cg-XII), as discussed above. • National commitments to open data policies.	Recommended

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 (see Wikipedia article on crowdsourcing) climate data by the NMHS.
	
	In summary, crowdsourcing is about obtaining data from a large pool of volunteers who would like to collaborate with the organization.
	
	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.
		Recommended

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. Examples of issues to consider include:	Recommended

	• 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 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?	

3.1.5.3
Commercial
	This component refers to policies that address and clarify issues on the use of climate-related data captured and maintained by commercial organizations.
	
	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?
	
	• 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 time limit for the use of the data? If so, what must be done with the 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
		Recommended

	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?	

3.1.6 Climatology policy

3.1.6.1 **Climate metadata**	This component refers to polices 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 **Data lineage traceability**	This component concerns policies that ensure that the CDMS is able to trace the lineage of climate data from published scientific texts and other papers back to raw observations. 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 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	Required

	Committee, 2011)		32

3.1.6.3 **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 a definition of a climatological day.
	
	• They should also cover the rules relating to the management of missing observations.
	
	Some inconsistencies have been found in a number of WMO guidelines. For example, conflicting approaches regarding the handling of missing observations when computing a daily or monthly average, and especially regarding the handling of a number of consecutive missing data, are presented in:
	
	• *Guide to Climatological Practices* (WMO-No.100)
	
	• *Calculation of Monthly and Annual 30-Year Standard Normals* (WMO/TD-No. 341), WCDP-10
	
	• *Handbook on CLIMAT and CLIMAT TEMP Reporting* (WMO/TD-No. 1188)
	
	Inconsistencies in WMO guides have also been noticed regarding the generation and storage of observations data recorded at minute frequency and the generation of hourly observation data. For example, the definition of climatological days may differ:
	
	• Between NMHSs
	
	• 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
		<mark>Required</mark>	

	– From 2346 LST the previous day to 2345 LST the current day – From 0900 LST the previous day to 0900 LST the current day There are also inconsistencies due to the use of either local standard time or daylight saving time. The same issues apply to the definition of a climatological hour.	
3.1.6.4 **Climate networks**	This component concerns a range of policies that determine the design of climatological networks and establish the station and network operations, including observation times, on-site quality control, observer training, station inspections, etc. For more information, see: • *Guide to Climatological Practices* (WMO-No. 100), section 2.5 The design of climatological networks, and section 2.6 Station and network operations • *Technical Regulations* (WMO-No. 49), Volume I, Part II, 1.3.1.1.2: The distribution of stations from which monthly surface climatological data are transmitted should be such that every 250 000 km^2 is represented by at least one station and up to 10 stations where the density of the regional basic synoptic network permits; the distribution of stations from which monthly upper-air climatological data are transmitted should be such that every 1 000 000 km^2 is represented by at least one station.	<mark>Required</mark>
3.1.6.5 **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. For more information, see: • *Manual on the Global Observing System* (WMO-No. 544), Volume I, Part III, 3.2.4: At reference climatological stations, any change in instrumentation should be such as not to decrease the degree of accuracy of any observations as compared with the earlier observations, and any such change should be preceded by an adequate overlap (at least two years) with the earlier instrumentation. • *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), Part II,	<mark>Required</mark>

	1.1.4 Climatological requirements: 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). • *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186), WCDMP-53	
3.1.6.6 **Quality assurance**	This component refers to policies that ensure that quality assurance issues within an organization are clearly understood. Some issues to consider are: • What level of quality assurance can the organization afford to maintain over the long term? • Will the organization conduct quality assurance 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? See: • *Technical Regulations* (WMO-No. 49), Volume IV	Recommended
3.1.6.7 **Future Climate Data Framework**	This component refers to policies that may be established by the future Climate Data Framework, as discussed in section 11.1. This may include, for example: • Common and consistent definitions of key datasets to be maintained by the NMHS, including common dataset names and service names. • A common definition of a climatological day, hour and so forth. • A consistent way of determining data	Recommended

	uncertainty.	
	• Common definitions of data quality.	
	• The establishment of a quality assessment classification for derived data.	
	• Authoritative taxonomies and code lists.	
	• Common policies for handling missing data when creating derived data.	
	• Common policies for deriving data when input data have differing levels of quality. For example, policies covering:	
	– The calculation of monthly averages when daily data are at different levels of quality.	
	– The generation of derived gridded datasets for a region where data for relevant stations are at different levels of quality.	
	• Common policies for all components discussed within this data policy section (3.1).	
	• The application of best practices for data homogenization techniques.	

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? • 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 users 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	**Required**

	the change 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?	**Required**
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**

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. For more information, see: • Wikipedia overview of IT service management • Information Technology Infrastructure Library (ITIL) website • ITIL publications	Recommended
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. • 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	Required

	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? • 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: • 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.	Recommended

	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 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. • CDMS component solutions could be developed that do not effectively use NMHS climate data, for example:	Recommended

		– 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. 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	
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**

4 Time-series climate data
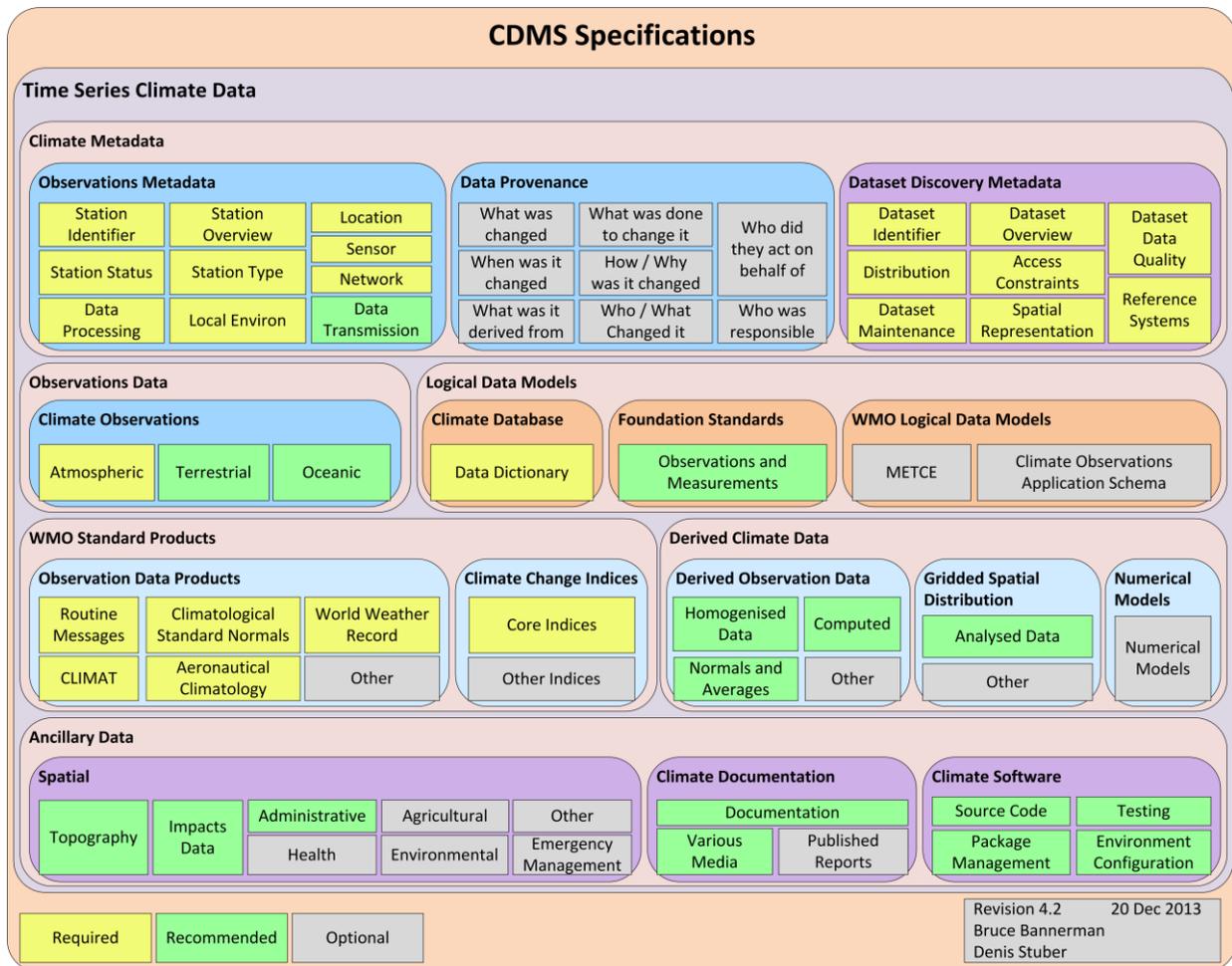
Figure 6. Time-series climate data

In this publication, the definition of climate data has been broadened beyond traditional meteorological data to ensure that consideration is given to a wide range of data and information relevant to climate and related uses. Considering modern demands for transparency and traceability, it is time to subject a broader set of climate data to more appropriate and consistent data management principles.

Therefore, the definition of climate-related data has been extended to include:

- Meteorological, hydrological, cryospheric and oceanographic data
- Climate metadata
- Observations of a wide range of variables, such as the GCOS ECVs
- The output generated by numerical climate models
- WMO standard products
- Data derived from terrestrial and remotely sensed observations
- Other related data

The management of climate data in particular presents challenges that are not currently found in many other data types. This is because:

- Each type of climate data is considered to be time-series data.
- It is becoming increasingly important to understand the provenance of the data.
- Climate data are considered as big data (see Wikipedia article on big data) which come in large volumes and in a variety of formats that make it difficult to manage the data effectively.

4.1   Observations data

4.1.1 Climate observations

This subsection refers to the system's capacity to handle climate observation variables. The list below is not comprehensive and additional variables may be required or recommended.

The list is currently based on the ECVs published in 2010 by GCOS. However, it is worth noting that there are other types of variables that are very relevant to climate observation, such as visibility, evaporation and meteorological phenomena.

Observations include measurements made by observers, traditional sensors and remote sensors.

For more information, see:
- GCOS website
- GCOS web page on ECVS

4.1.1.1
 Atmospheric
	Observations relevant to the GCOS ECVs are needed to support the work of the United Nations Framework Convention on Climate Change and the Intergovernmental Panel on Climate Change.
	
	 Within this component, observations relating to the following atmospheric ECVs (over land, sea and ice) are either required or recommended, as shown below:
	
	 Required
	 • Surface
	    – Air temperature
	    – Wind speed and direction
	    – Water vapour
	    – Pressure
	    – Precipitation
	    – Surface radiation budget
	 • Upper air
	    – Air temperature
	    – Wind speed and direction
	    – Water vapour
		Required

	Recommended
	
	• Upper air
	   – Cloud properties
	   – Earth radiation budget (including solar irradiance)
	• Composition
	   – Carbon dioxide
	   – Methane
	   – Ozone
	   – Aerosol
	   – Other
	For more information, see:
	• GCOS web page on ECVs
	• *Guide to Climatological Practices* (WMO-No. 100), Table 2.1	

4.1.1.2 **Terrestrial**	Terrestrial ECVs: • River discharge • Water use • Groundwater • Lakes • Snow cover • Glaciers and ice caps • Ice sheets • Permafrost • Albedo • Land cover (including vegetation type) • Fraction of absorbed photosynthetically active radiation • Leaf area index • Above-ground biomass • Soil carbon • Fire disturbance • Soil moisture For more information, see: • GCOS web page on ECVs • *Guide to Climatological Practices* (WMO-No. 100), Table 2.1	Recommended

4.1.1.3
Oceanic
	Oceanic ECVs:
	• Surface
	   – Sea-surface temperature
	   – Sea-surface salinity
	   – Sea level
		Recommended

	– Sea state
	– Sea ice
	– Surface current
	– Ocean colour
	– Carbon dioxide partial pressure
	– Ocean acidity
	– Phytoplankton
	• Subsurface
	   – Temperature
	   – Salinity
	   – Current
	   – Nutrients
	   – Carbon dioxide partial pressure
	   – Ocean acidity
	   – Oxygen
	   – Tracers
	For more information, see:
	• GCOS web page on ECVs
	• *Guide to Climatological Practices* (WMO-No. 100), Table 2.1	

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 describes the content and structure of information resources at an abstract level. When implemented in a CDMS, the logical data model underpins the design of the information objects managed by the system and helps to determine the questions that one may ask of the system when querying those information objects and their interrelationships. Furthermore, the logical data model may be used as the basis for developing data formats within which the information objects can be serialized for exchange between systems – thus ensuring interoperability between those systems.

WMO is currently developing a logical data model termed the Modèle pour l' Échange des informations sur le Temps, le Climat et l'Eau[2] (METCE), which is intended to provide a basis for application-specific data models, including those used for climate observing and climate data management.

Ideally, the underlying database structure will be based on the logical data model.

---

[2] Information model for weather, water and climate; also known as the METeorology Community Exchange model.

4.2.1 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.

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 metadata) and integrating them with observations data.
- 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 to homogenized data.
- Managing spatial and time-series data.

Recommended

- Covering at least the GCOS ECVs.
- Using a robust data model that takes into account the requirements of open spatial standards, particularly the ISO 19156:2011 *Geographic information – Observations and measurements* standard, METCE and the WMO climate observations application schema (see component 4.2.3.2).
- Managing metadata related to data provenance. This entails ensuring that each change to an observation is recorded for future recovery, and 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?
    - Ensuring that the reason for each observation change is recorded.
- 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, 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 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 models • Providing support for the information management concepts of semantics and linked data		
4.2.1.1 **Data dictionary**	This component represents the data dictionary, which describes the database structure, data model and other elements used by the climate database.	**Required**

4.2.2 Foundation standards

4.2.2.1 **Observations and measurements**	This component represents technology that provides rules and a standardized approach for modelling observations data, regardless of the domain. In essence, the ISO 19156:2011 *Geographic information – Observations and measurements* 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 WaterML and the Meteorological Information Exchange Model of the International Civil Aviation Organization (ICAO). It also underpins current work on the WMO logical data model called METCE (see below). For more information, see: • ISO 19156:2011, *Geographic information – Observations and measurements* • OGC Abstract Specification: Geographic information – Observations and measurements	Recommended

4.2.3 WMO logical data models

4.2.3.1 **METCE**	The METCE component represents technology that provides rules and a standardized approach for modelling observations and simulations in the weather, water and climate domains. METCE is an application schema conforming to ISO 19109:2005 *Geographic information – Rules for application schema*. Furthermore, METCE is a profile of the Observations and measurements	Optional

	standard that provides domain- and application-specific semantics for use within the weather, water and climate domains.	
	The initial iteration of METCE and its companion model, the Observable Property Model, were developed by the Task Team on Aviation XML to support the ICAO Meteorological Information Exchange Model.	
	However, METCE will provide a common semantic basis for a growing number of data products relating to observation and simulation within WMO. Not only will this simplify the requirements for software systems working with WMO products, it is also expected to simplify the mappings between WMO data products and counterparts from other communities such as CF-netCDF.	
	As at December 2013, plans have been made to provide mappings/rules to convert from the METCE application schema to BUFR sequences and/or GRIB templates at some point in the future.	
	For more information, please see:	
	• AvXML-1.0 data model	
	• ISO 19109:2005, *Geographic information – Rules for application schema*	
4.2.3.2 Climate observations application schema	This component represents technology that provides rules and a standardized approach for modelling climate observations data.	
	It is anticipated that METCE will be used as the basis for developing an application schema that will provide more detailed semantics and constraints specific to a given domain or application. In this way, METCE will provide the basis for an application schema developed to support the wide array of climate observation applications.	Optional
	The scope of such an application schema is expected to cover both the climate observations themselves and the associated observation metadata (see subsection on observations metadata (4.3.1)).	

4.3　　Climate metadata

The term climate metadata is defined in this publication as the suite of supporting data required to effectively manage climate data and assess the data's fitness for purpose.

Climate metadata are made up of the following components:

- Observations metadata   Time-series data that describe how, when and where meteorological observations were made and the conditions they were made under.

- Discovery metadata   Information intended to facilitate the discovery and assessment of a dataset to determine if it is fit for reuse for a purpose that may be at odds with the reason for which it was originally created.

- Data provenance metadata   Information relevant to climate data that allows end-users, including data managers, scientists and the general public, to develop trust in the integrity of the climate data.

The following section expands on each component of climate metadata.

4.3.1 Observations metadata

This subsection covers access to and management of station metadata and platform metadata. Station and platform metadata are time-series data that describe how, when and where meteorological observations were made and the conditions they were made under. They are 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. This type of metadata is referred to as observations metadata in this publication.

It is anticipated that application schemas (also known as logical data models) will be developed to formally define the structure and content of the information required to describe climate observing stations, sensors and platforms (see the Climate observations application schema component (4.2.3.2))

Note:
- As a general rule, it will be necessary to record and maintain the details of any change to observations metadata in order to understand the context surrounding specific climate data and to support future data homogenization activities. In addition to details of the change, specific reference must be made to:
    - The date/time of the change.
        - Note: It may not always be possible to define the exact date of the change, for example when a change happens between two site visits. Therefore, it may be more appropriate to include a period of time during which the change occurred.
    - The reason the change was made.
    - The beginning and end dates of the prior value.
    - Any date/time reference will need to be constrained by the appropriate temporal datum to ensure that date handling is consistently applied.

For more information, see:
- *Guide to Climatological Practices* (WMO-No. 100)

- *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8)
- *Guide to the Global Observing System* (WMO-No. 488), Appendix III.3 Automatic weather station metadata
- *Manual on the Global Observing System* (WMO-No. 544), Volume I, Attachment III.1 Standard set of metadata elements for automatic weather station installations
- Discussion paper on stations metadata and the WMO Core Profile (Bannerman, 2012)
- Draft paper on the climatological needs for minimum stations metadata in the frame of WMO publication No. 9, Volume A (Stuber, 2012)
- *Guidelines on Climate Metadata and Homogenization* (WMO-No. 1186), WCDMP-53

4.3.1.1 **Station identifier**	This component supports the management of identifiers associated with the observation station or platform. Identifiers include: • A globally unique WMO identifier. The use of this identifier must become a priority in order to support future global analysis. See recommendation 11.6 of this publication. • Other identifiers or aliases used for the station. • A history of past used identifiers, including historical WMO identifiers. • The beginning and end dates of each historical identifier used for the station.	**Required**

4.3.1.2
Station overview
	This component covers what to provide in an overview of the observation station or platform.
	This should include:
	• Station owner
	    – If required, the sensor owner
	• Station manager
	    – If required, the sensor manager
	• Maintenance authority
	    – If required, the sensor maintenance authority
	• Station licence agreement
	    – If required, the sensor licence agreement
	• Station data usage constraints
	    – If required, the sensor data usage constraints
	• Purpose of the station
	• Observation practices
	• Observation schedule – this is particularly relevant for stations that use manual observation methods and where observations are not taken on a continuous basis
	• Definition of which datasets provide the actual
		Required

	observations data for a given station, sensor and phenomenon combination, together with the URL of the relevant discovery metadata records • Observers and maintenance personnel, including their names, experience and training level • Station logistics, including consumables, electricity suppliers, communications suppliers, etc.	
4.3.1.3 **Station status**	This component supports recording the period(s) of activity during which observations were being made at the station or platform. As it is possible for stations to close and then reopen at a later time, the time period of each status is also required. Valid operational status codes are: • Operational • Not operational	**Required**
4.3.1.4 **Station type**	This component supports recording the type of station or observation platform. Ideally, the type will be recorded for each instrument used at that station. At a minimum, the station type is to be recorded in accordance with the following guidelines: • *Manual on Codes* (WMO-No. 306), Volume I.1, code 1860, and Volume I.2, code 0 02 001 In addition, there may be multiple definitions of station types used by the NMHS and other organizations.	**Required**
4.3.1.5 **Location**	This component covers the recording of details relating to the location of the station or observation platform. As discussed in the Sensor component below (4.3.1.7), recording the location of each sensor at the station is also mandatory. The following information is required: • Latitude • Longitude • Elevation • Spatial reference system (horizontal and vertical) • Date/time of the survey observation used to record the location of the station and/or sensor • Temporal reference system • Method used to determine the location of the	**Required**

	station	
	• Positional accuracy of location	
	• Date/time the station or sensor moved, together with previous locations	
	• Administrative boundaries within which the station is located (as required by the NMHS)	
	• Time zone	
	In exceptional circumstances, it may be necessary to move a station and keep the same station identifier. This should only be done in accordance with future global climate data policies and data governance processes (see recommendation 11.1 of this publication). Typically, this will involve parallel observations at the old and new station over a period of time (two years, for example).	
	If this is necessary, the time of the move should be recorded, together with current and past location details.	
	The precision required for the latitude, longitude and elevation should be in accordance with guidance provided by the Commission for Instruments and Methods of Observation.	
	While not authoritative, the final report of the first session of the Commission for Instruments and Methods of Observation Expert Team on Standardization suggests that the precision required for latitude and longitude measurements is one second (of arc), which equates to approximately 30 meters at the equator. This degree of precision should be achievable using a survey observation process that uses handheld GPS techniques.	
	Administrative boundaries may refer to different types of boundaries that contain the station, including:	
	• Political boundaries, such as state, regional or district boundaries	
	• Administrative boundaries, such as the forecast district	
	• Natural boundaries, such as hydrological, catchment or topographic areas	
	For more information, see:	
	• ISO 19111-2:2009 *Geographic information–Spatial referencing by coordinates, Part 2: Extension for parametric values*	
	• *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), section 1.3.3.2 Coordinates of the station. Specifically	

	note the instructions on determining the elevation of a station relative to the raingauge. • OGC Abstract Specification: Spatial referencing by coordinates	
4.3.1.6 **Local environment**	This component concerns the recording of information on the local environment surrounding the station or observation platform. The following information is required: • Site location diagram • Site plans – For an example, see *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), Figure 1.1 • Site skyline diagram • Site photographs and video showing the surroundings and instrument layout • Station exposure • Site roughness • Type of soil • Type of vegetation • Surrounding land use • Date/time of each visit For more information, see: • *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), section 1.3.3 Siting and exposure, Annex 1.B Siting classifications for surface observing stations on land, and Annex 1.C Station exposure description	<mark>Required</mark>

4.3.1.7
Sensor
	This component covers the recording of details relating to the meteorological sensors and/or instruments used at the station or observation platform. In this publication, the term sensor will be used to cover all instrument types.
	The following information is required:
	• Sensor description, including:
	  – Name
	  – Type
	  – Serial number
	  – Brand and model details
	  – Photograph of sensor in situ
	  – Supplier
	  – Manufacturer
	  – Location of manuals
		<mark>Required</mark>

	- Sensor firmware, version and dates during which each version was used - Length of time the observation data are stored locally on the sensor, prior to deletion - Sensor installation details, including: - Technician and organization that installed the sensor - Date sensor was installed - Sensor status, including: - Operational status - Operational - Not operational - Defective - Testing - Date/times applicable for each status - Sensor maintenance: - Scheduled maintenance - Actual maintenance - Result - Replacement of consumables - Sensor uncertainty: - System performance statistics claimed by manufacturer - Sensor calibration results - Observed sensor performance characteristics - Sensor siting details: - Instrument height above ground - Station exposure description - As discussed in the Location component (4.3.1.5), recording the location of each sensor is required. - Recommended sensor settings for optimal operations on site - Details of what meteorological variable is being observed by the sensor (i.e. the observed property), including: - Phenomena observed - Frequency of measurement - Frequency of acquisition - Units of measurement - Precision of measurement	

4.3.1.8
 	
Data processing
	This component concerns the recording of details relating to any data processing that has occurred to convert a sensor's signal into its recorded observation value.
	
	The following information is required:
	• Software, including:
	   – Version
	   – Software language
	   – Software name
	   – Location of software source code
	   – Description of processing applied (for example, whether values were calculated per minute, hour or other)
	   – Formula/algorithm implemented
	   – Processor details (the version, type of central processing unit and so forth)
	   – Date/time covering the period of validity of the method
	• Input source (instrument, element and so forth)
	• Data output, including:
	   – Data format and version of format
		Required

4.3.1.9
 	
Data transmission
	This component refers to the recording of details relating to the transmission of data from stations or observation platforms.
	
	The following information is required:
	• Sensor communications, including:
	   – Frequency of transmission
	   – Time of transmission
	   – Primary communication details
	   – Redundant communication details
	   – Network addresses
	   – Method of transmission
	
	Note: Some NMHSs with more advanced IT infrastructures may choose to store this type of information within their configuration management system. In these instances, it is important to ensure that at least the frequency and time of the transmission are replicated in the observations metadata.
		Recommended

4.3.1.10 **Network**	This component concerns the recording of details relating to the observation network(s) that stations or observation platforms may belong to. The following information is required: • Network name (such as Regional Basic	**Required**

	Climatological Network, Regional Basic Synoptic Network, GCOS, GCOS Upper-Air Network or National Climate Network)	

	• Network priority:
	    – Critical
	    – Essential
	    – Not applicable
	• Time of observations
	• Reporting frequency
	• Date/time of network membership
	• There is a possibility that a station does not belong to a network. This information is also useful.	

4.3.2 Dataset discovery metadata

This subsection 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 spatial dataset to determine if it is fit for reuse for a purpose that may be at odds with the reason 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.

Note that some of the components below may be in addition to the WMO Core Profile of the ISO 19115 *Geographic information – Metadata* standard. This is not expected to be an issue as the WMO Core Profile does not restrict the use of additional ISO 19115 records.

For more information, see:

- Discussion paper on stations metadata and the WMO Core Profile (Bannerman, 2012)
- *WMO Core Metadata Profile version 1.3*, Part 1 Conformance requirements
- *WMO Core Metadata Profile version 1.3*, Part 2 Abstract test suite, data dictionary and code lists
- Tandy (2010) also provides a useful introduction to discovery metadata. However, note that the specifications part of this text has been superseded.
- ISO 19115 *Geographic information – Metadata*

4.3.2.1 **Dataset identifier**	This component represents a unique identifier used to identify the dataset.	<mark>Required</mark>
4.3.2.2 **Dataset overview**	This component gives an 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.	<mark>Required</mark>
4.3.2.3 **Dataset data**	This component represents a general assessment of the quality of a dataset.	<mark>Required</mark>

quality		
4.3.2.4 **Distribution**	This component covers information about the distributor of and options for obtaining a dataset.	<mark>Required</mark>
4.3.2.5 **Access constraints**	This component provides information on the restrictions in place for a dataset.	<mark>Required</mark>
4.3.2.6 **Dataset maintenance**	This component provides information on the scope and frequency of updates and maintenance conducted on a dataset.	<mark>Required</mark>
4.3.2.7 **Spatial representation**	This component covers information on the mechanisms used to represent spatial information within a dataset.	<mark>Required</mark>
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.	<mark>Required</mark>

4.3.3 Data provenance

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

Data provenance allows an end-user to understand the history of each piece of data, and thus 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 a number of 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.

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 more than 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. (W3C, 2013*a*)

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

The concepts presented here for climate data provenance are quite embryonic and need further work to ensure that they can be implemented effectively. See the related recommendation in Chapter 11.

Maintaining a dataset with high levels of data provenance metadata is expected to be quite expensive and, as a result, will be limited to data of high importance 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 could perhaps be included as a policy within a future climate data framework.

For more information, see:
- Overview of the PROV family of documents (W3C, 2013*b*)
- PROV data model specification (W3C, 2013*a*)

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.	Optional
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.	Optional
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).	Optional
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	Optional

	• Details describing what was done	
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	Optional
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.	Optional
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 is identified.	Optional
4.3.3.8 **Who was responsible?**	This component refers to the processes, software and governance arrangements that ensure that the person or role that authorized the change is identified.	Optional

4.4   WMO standard products

4.4.1 Observation data products

This subsection outlines the types of data products that NMHSs have committed to generate and provide to WMO.

4.4.1.1 **Routine messages**	This component represents data computed from observation data for use in WMO products. An example is the daily minimum and maximum temperature, evaporation and evapotranspiration that are typically transmitted via the Global Telecommunication System (GTS) as a SYNOP message or in the corresponding table-driven code form (TDCF) for SYNOP. For more information, see: • *Manual on Codes* (WMO-No. 306)	**Required**
4.4.1.2 **Climatological**	This component covers monthly and annual standard normals. See Wright (2012*a*) for a description of a	**Required**

standard normals	recommended change in the method of calculating climatological standard normals. The approach is understood to comprise the following. (However, as at December 2013 this has yet to be officially endorsed.) • A fixed reference period (1961–1990) for long-term climate variability and change assessment. This is to be adopted as a stable WMO reference period, until such time as there is a compelling scientific case for changing it. • A varying 30-year period updated every 10 years (suitable for most climate services). The current period is 1981–2010. For more information, see: • *Calculation of Monthly and Annual 30-Year Standard Normals* (WMO/TD-No. 341), WCDP-10 • *Technical Regulations* (WMO-No. 49), Volume II • *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) • *A Note on Climatological Normals: Report of a Working Group of the Commission for Climatology*, Technical Note No. 84 • *The Role of Climatological Normals in a Changing Climate* (WMO/TD-No. 1377), WCDMP-61 • Discussion paper on the calculation of the standard climate normals (Wright, 2012*a*)	
4.4.1.3 **CLIMAT**	This component concerns CLIMAT messages in either traditional alphanumeric codes (TAC) or TDCF formats. These messages are transmitted to WMO via the GTS. Note: The use of TAC is being phased out. 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, Part C, section d. Regulations for reporting traditional observation data in table-driven code forms (TDCF): BUFR or CREX	**Required**

4.4.1.4 **World Weather Records**	This component covers the annual World Weather Records. For more details, see: • *Guidelines on Submission of the World Weather Records Tenth Series (2001–2010),* WCDMP-77 • World Weather Records website	<mark>Required</mark>
4.4.1.5 **Aeronautical climatology**	This component refers to the monthly aerodrome climatology summary – tabular forms (models A to E). For more explanation, see: • *Technical Regulations* (WMO-No. 49), Volume II, C.3.2 Aeronautical climatology	<mark>Required</mark>
4.4.1.6 **Other**	This component covers other WMO standard products, as may be required. Other standard products are to be confirmed, particularly for the hydrological, agricultural and marine domains.	Optional

4.4.2 Climate change indices

This subsection represents the recommendations of the Commission for Climatology in the elaboration of climate change indices. The Commission for Climatology/Climate Variability and Predictability (CCl/CLIVAR) Working Group on Climate Change Detection has been coordinating an international effort to develop, calculate and analyse a suite of 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.

Those indices have been split in two categories: core indices and approved indices.

For more information, see the website of the Joint CCl/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI).

4.4.2.1 **Core indices**	This component represents the ETCCDI core climate change indices. As at June 2013, 27 core indices have been defined, see: • ETCCDI website	<mark>Required</mark>
4.4.2.2 **Other indices**	This component covers other climate change indices. As at June 2013, different research groups have defined different indices for their particular purposes. One example is the Statistical and Regional dynamical Downscaling of Extremes for European regions (STARDEX) project, mentioned in the ETCCDI website referred to above.	Optional

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.

4.5.1.1 **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
4.5.1.2 **Computed**	This component deals with derived data computed from observations for NMHS products. The computation shall be in accordance with the climatology policies in place. Some examples are: • Data generation of accumulation, 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. • The generation of any statistical parameter required for products, such as extreme value analysis, homogenized data and others. • Climate indices such as computed teleconnection indices.	Recommended
4.5.1.3 **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 (daily, hourly, 5 days, 10 days, monthly and so forth) • Period averages for any period (for example 5, 10, 30 or 100 years) For more information, see: • *The Role of Climatological Normals in a Changing Climate* (WMO/TD-No. 1377), WCDMP-61	Recommended
4.5.1.4 **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 subsection 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), section 5.9 Estimating data.

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

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.

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

4.5.3.1 **Numerical models**	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. • Reanalysis – designed for 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. This unchanging framework provides a dynamically consistent estimate of the climate state at each time step. Some available reanalysis products include ERA-40 (40 years) and ERA-Interim (1979 to the present) from the European Centre for Medium-Range Weather Forecasts, and the Twentieth Century Reanalysis project (1871–2011) from the National Oceanic and Atmospheric Administration, United States. • Numerical weather prediction See also: • *Guide to Climatological Practices* (WMO-No. 100), section 6.7 Climate models and climate outlooks	Optional

4.6   Ancillary data

This section covers data required to support CDMSs.

4.6.1 Spatial

This subsection represents a wide range of spatial information typically used to provide context to climate data or as an input for spatial analysis processes. These data may be presented in vector, image, gridded or multidimensional array data formats.

Typically, spatial representations of data contain aspatial attributes that describe various properties of spatial features. The spatial and aspatial attributes of the data can be used to support a variety of spatial analysis processes.

The components listed below are indicative of the types of spatial data that could be relevant to climate. This list is not exhaustive.

See also:

- Statement of guidance for climate, Attachment 1 Requirements for climate data (Wright, 2012*b*)

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	Recommended
4.6.1.2 **Emergency management**	This component concerns datasets that are useful for supporting emergency management and related warning systems.	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.	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.	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 centres. • Data relating to the distribution of particular flora and fauna, etc.	Optional

4.6.1.6 **Administrative**	This component covers administrative data. Some examples are: • Localities and gazetteers • Administrative boundaries • Transportation networks • Cadastres	Recommended
4.6.1.7 **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, etc. • Infrastructure damage caused by a range of events such as floods, bushfires or cyclones. • Changing land use, such as agricultural adaptations due to a changing climate.	Recommended
4.6.1.8 **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: • Peer-reviewed papers • Climate change impact studies • Climate statements and studies • Assessments from the Intergovernmental Panel on Climate Change • Monthly and annual summaries As this is essentially a scientific and intellectual process, this component will not be expanded upon in this publication.	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: • CDMS technical and user documentation • Text on a web page • Diagrams representing climate processes, such as the Community Earth System Model	Recommended

	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 • NMHS policies and practices • Various documents and reports • Training documentation • 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

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.

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.

4.6.3.1 **Source code**	This component deals with managing the source code of the software used to process climate data. This component should have the following	Recommended

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

5 Climate data management
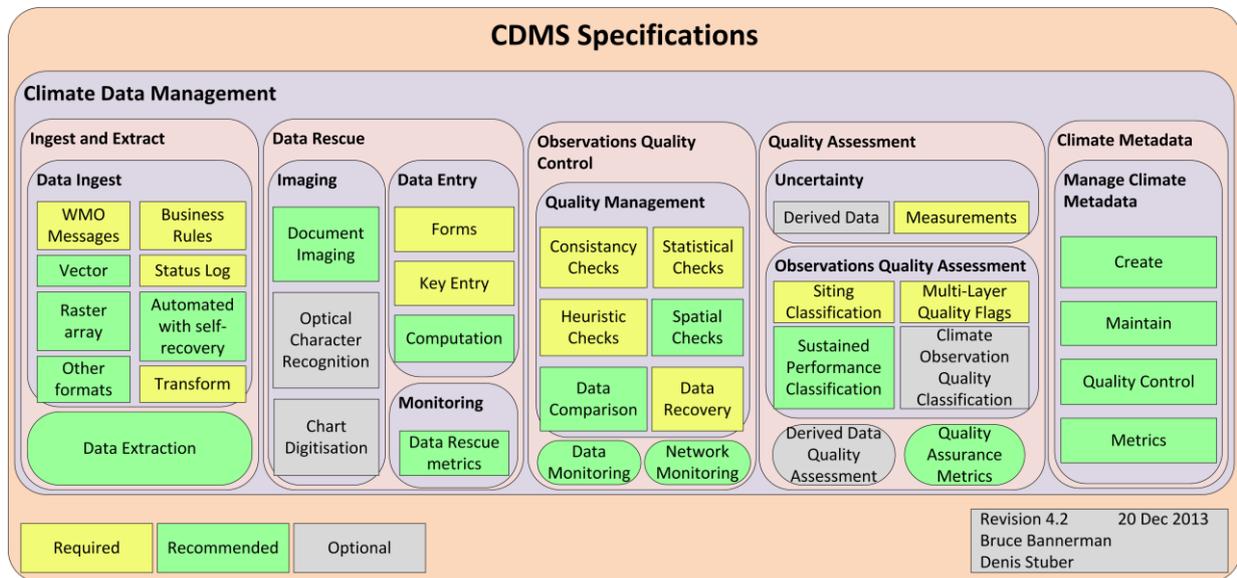
Figure 7. Climate data management

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.

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

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: • Action required when new phenomena 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	**Required**

	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 accepted temperature is 60°C. • 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.	

5.1.1.2
WMO messages
	This component allows for the import of data from a range of WMO message formats, including TAC and 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
	    – GRIB
	
	• Alphanumeric:
	    – CREX
	    – SYNOP
	    – TEMP
	    – SHIP
	    – METAR
	    – World Weather Records
	
	Note: While the use of TAC formats are being phased out, support for them will still be required
		<mark>Required</mark>

	by this component to support the ingest of historical data. For more information, see: • WMO international codes	
5.1.1.3 **Vector**	This component supports the import of a series of vector spatial formats. For example: • Shapefile • Geography Markup Language (GML) (see OGC GML web page)	Recommended
5.1.1.4 **Raster array**	This component supports the import of a series of raster array spatial formats. For example: • CF-netCDF • Hierarchical data format • ArcInfo ASCII • GeoTIFF	Recommended
5.1.1.5 **Other formats**	This component covers the import of a range of other formats. For example: • Photographs (PNG, JPEG, TIFF, etc.) • Scanned documents • PDF files • ASCII generic formats such as CSV • Data managed in spreadsheets • Tabular formats, such as the import of data from a relational database management system	Recommended
5.1.1.6 **Status log**	This component concerns the recording of each ingest activity status in order to: • Monitor the ingest job status. • Automatically recover failed ingests. • Record warning and other error messages to enable manual intervention if required, for example if expected data are not received.	**Required**
5.1.1.7 **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,	Recommended

	including: • Corrupted messages • Network failures • Hard disk failures • Database failures • Upstream data flow disruptions	
5.1.1.8 **Transformation**	This component supports the transformation of an ingest record. This may include: • Transforming data from one format to another. • Transforming codes into formats more suitable for the destination climate database. • Correcting records that have been abbreviated in accordance with accepted local observation practice.	<mark>Required</mark>

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.	Recommended

5.2   Data rescue

5.2.1 Imaging

5.2.1.1 **Document**	This component supports the functionality required to digitally capture a physical document and store the resultant file and associated discovery metadata, perhaps within the climate	Recommended

imaging	database. Some examples of the types of documents to be digitally captured are: • Scanned paper observation forms • Scanned microfiche/microfilm • Relevant observations metadata documents such as instrument calibration reports • Technical manuals • Site location plans and sections For more information, see: • *Guidelines on Climate Data Rescue* (WMO/TD-No. 1210), WCDMP-55	
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.	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 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. • Extract numeric points from the chart. • Calculate a value for those points. • Store the resultant data in the climate database.	Optional

5.2.2 Monitoring

5.2.2.1 **Data rescue metrics**	This component maintains metrics relating to the capture of historical observations data. These may contain: • Name and brief description of data rescue project • Countries where activity is taking place • Contact person for project • Types of data rescued • Summary and per cent digitized • Summary and per cent scanned	Recommended

	• Summary and per cent scanned but not digitized • Summary and per cent undigitized	

5.2.3 Data 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.
- 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 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.
- 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. Most organizations will not have suitable controls in place. Therefore, key-in via the Internet should be avoided as a general rule.
- NMHS data policy should provide guidelines as to appropriate data quality considerations applied to data that are manually entered.

5.2.3.1 **Forms**	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). • Efficient and effective data entry that minimizes operator fatigue and automatically calculates appropriate values.	**Required**

	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.	<mark style="background-color: yellow">**Required**</mark>
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 dew point have been entered. • Decoding shorthand codes and replacing them with appropriate values.	<mark style="background-color: lime">Recommended</mark>

5.3　　Observations quality control

5.3.1 Quality management

For more information, see:

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

• *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

• *Guidelines on the Quality Control of Surface Climatological Data* (WMO/TD-No. 111), WCP-85

• *Guidelines on Climate Data Management* (WMO/TD-No. 1376), WCDMP-60

• *Guide on the Global Data-processing System* (WMO-No. 305), Chapter 6 Quality control procedures

5.3.1.1 **Consistency checks**	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 as discussed in the *Guide to Climatological Practices* (WMO-No. 100), section 3.4.6 Consistency tests. 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?	Required
5.3.1.2 **Data comparison**	This component covers a series of tests that use and cross-reference data from a number of sources to validate suspect observations. Some examples of datasets that may be cross-referenced are: • Observations data showing daily precipitation at a station • Radar data covering the station • Synoptic forecast charts • Satellite imagery	Recommended
5.3.1.3 **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. Some examples are problems typically caused by: • Inexperienced operators. • Instruments that are not or are incorrectly calibrated. • Operator behaviour or organizational policy, for example not recording rainfall data over a weekend period and aggregating the results on the following Monday. • 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	Required

	the site location, such as a new building structure or trees obstructing the flow of the wind in that direction.	
5.3.1.4 **Statistical checks**	This component covers a number of tests that statistically analyse historical data to detect 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: • Climate tests that highlight extreme climatic values, such as a record maximum air temperature. • Flatline tests where a constant value exceeds the 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 50 degrees colder than all others during the day. • Rapid change tests conducted in a time series to identify 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 for all following observations.	**Required**
5.3.1.5 **Spatial checks**	This component covers a range of spatial tests to detect 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 or similar analysis to derive spatial patterns against which anomalous and possibly erroneous station values stand out.	Recommended
5.3.1.6 **Data recovery**	This component refers to the processes, policies, governance arrangements, audit processes, etc., that enable the recovery and insertion of data in the climate database, possibly overwriting existing data. This component involves a number of manual processes undertaken by experienced and well-trained personnel, supported by effective	**Required**

	technology, governance and data management processes, to investigate anomalous observations and either accept or reject suspect records. Personnel will typically review and consider a wide range of data in their investigations, such as raw records, synoptic charts, satellite imagery, radar and other types.	

5.3.2 Network monitoring

	This component keeps track of data received from an observational network in order to monitor network performance and potentially detect and alert data managers of possible problems. Some examples are: • Statistics showing the current state of data ingests 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 systematic error such as an undetected software error. • Records received in advance of the observation time may highlight training issues. • 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	Recommended
5.3.2.1 **Network monitoring**		

5.3.3 Data monitoring

	This component monitors ingested and derived observation data to detect and resolve potential systemic issues. 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: • Daily, weekly, monthly, seasonal and annual summaries for a range of observation variables.	Recommended
5.3.3.1 **Data monitoring**		

	• 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.	
	Another example is the generation of a number of metrics to assist with climate data management activities.	
	This component maintains metrics relating to the capture of historical observations data, for example:	
	• Summary and per cent of data that have undergone quality control.	
	• Percentage of data at each level of quality control.	

5.4　Quality assessment

5.4.1 Observations quality assessment

This subsection 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 Meteorological Instruments and Methods of Observation* (WMO-No. 8), Annex 1.B Siting classifications for surface observing stations on land. | **Required** |
| 5.4.1.2

Sustained performance classification | This component refers to the processes, software, governance mechanisms and analysis that classify sensors according to their sustained performance over time.

The best description found to date on how to determine this classification may be found in Annex III of the final report of the first session of the Commission for Instruments and Methods of Observation Expert Team on Standardization. | Recommended |

		Note: A more objective approach to developing this classification for the global WMO community is 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 a specific record of data. This will facilitate: • Future analysis that requires data of a specific quality flag value. • 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), pp. 3–8 to 3–9. This reference describes a way of flagging quality based on a combination of: • Data type (original, corrected, reconstructed or calculated) • Validation stage • Acquisition method This approach is still quite limited. It does not provide a clear way of determining just what level of quality control a record has been subjected to. 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 (without providing actual details). If a record has passed all such tests, can it be considered to be better quality than one that has not passed any test? 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. It is becoming increasingly apparent that organizations will need to retain observations at multiple levels of quality from the raw observation through various edit 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	==Required==

	this classification for the global WMO community is required.	
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 • Sustained performance classification • Regular maintenance and calibration of sensor • Sensor reliability • Uncertainty inherent in observations • Observation quality control processes • Multilayer quality flags • Lineage • Homogeneity • Other appropriate factors See also the summary of findings of the seventh Data Management Workshop of the European Climate Support Network (ECSN) held at the Danish Meteorological Institute, in particular: Noting that "everybody" talks about different levels of Quality Control [QC] and (almost) nobody uses the same wording or nomenclature – it is recommended that an **overview of QC nomenclature in ECSN is worked out.** It might be considered if such an overview could form the basis for a recommended set of QC wordings. (Kern-Hansen, 2009)	Optional

5.4.2 Derived-data quality assessment

5.4.2.1 **Derived-data quality**	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	Optional

assessment	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.	

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 network managers and climate data specialists to validate the performance of quality assurance software and processes. This can be done, for example, by reviewing automatically generated reports that: • Summarize observational errors detected by each quality assurance test. • Summarize false positives and valid errors detected. • Compare the performance of current quality assurance metrics with historical averages. These types of metrics can also help data and network managers improve quality assurance processes and software.	Recommended

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 (p. 13), all observations have an element of uncertainty:

> The observation error typically has a systematic 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 errors in a property value are important in the context of a data analysis or processing application, then the details of the act of observation which provided the estimate of the value are required.

This functionality will support:

- Future statistical analysis that takes into account the uncertainty inherent in data.
- Communication of data uncertainty.

For more information, see Wikipedia articles on:

- Uncertain data
- Uncertainty

5.4.4.1 **Measurements**	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. The *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8) provides a number of examples per meteorological variable. For more information, see: • *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), Annex 1.D Operational measurement uncertainty requirements and instrument performance • Annex III of the final report of the first session of the Commission for Instruments and Methods of Observation Expert Team on Standardization	**Required**
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 gridded data that have been derived from observation data. Many factors can contribute to the uncertainty inherent in gridded derived data. Some examples are: • Uncertainty inherent in the source observations data. • Uncertainty inherent in the location of sensors/stations used to generate the grids. • The relative accuracy of the algorithms used	Optional

	to generate the derived data.	
	• The precision of variable data types used in the software that generates derived data. It is also worth noting that a number of these factors may propagate through the data derivation process.	

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, 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. While the Create and Maintain components (5.5.1.1 and 5.5.1.2, respectively) are classified as recommended, in reality data provenance metadata has not yet been adequately defined. Therefore, a pragmatic approach would rightly not address the creation and maintenance of data provenance metadata until this has been rectified. The creation and maintenance of discovery and observations metadata, however, is required.

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.	Recommended
5.5.1.2 **Maintain**	This component covers the processes, software and governance mechanisms required to effectively and efficiently maintain climate metadata.	Recommended
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.	Recommended
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? • Which datasets do not have discovery metadata records?	Recommended

	More work is required to provide effective guidance on this component.	

6 Climate data analysis
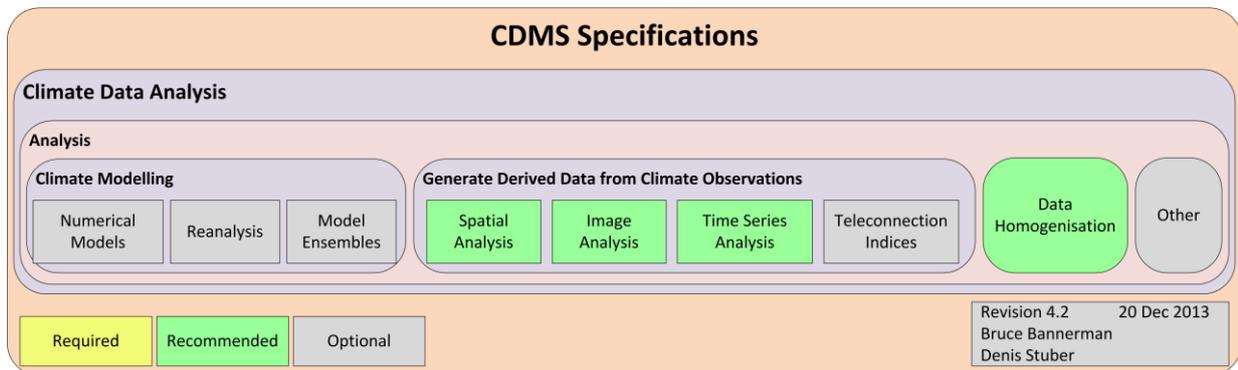
Figure 8. Climate data analysis

The Climate data analysis component represents the range of functionalities typically required to effectively analyse climate data.

6.1 Analysis

This section describes a series of processes used to analyse climate data.

More work will be required to expand on this section 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 mechanisms that provide numerical models such as general circulation models (GCMs), also known as global climate models. A Wikipedia article defines general circulation models as: [A] mathematical model of the general circulation of a planetary atmosphere or ocean and based on the Navier–Stokes equations on a rotating sphere with thermodynamic terms for various energy sources (radiation, latent heat). These equations are the basis for complex computer programs commonly used for simulating the atmosphere or ocean of the Earth. Atmospheric and oceanic GCMs (AGCM and OGCM) are key components of **global climate models** along with sea ice and land-surface components. GCMs and global climate models are widely applied for weather forecasting, understanding the climate, and projecting climate change.	Optional

	Such climate numerical models could be global in scale (like GCMs) but they could also be regional, with generally a higher precision. Downscaling techniques are often used when creating regional models. Models can be based on: • The rules of physics, biology and chemistry • Statistical rules • A mix of dynamic and statistic methods The use of climate models includes: • Simulation of the present climate • Simulation of the future climate • Palaeoclimate reconstruction • Seasonal forecasts • Decadal prediction Note: Most NMHSs would not have the resources needed to effectively manage the infrastructure and software required to support this component. However, the data output from such components may be useful and should be available for NMHSs via a number of sources, including Regional Climate Centres. For more information, see: • Wikipedia article on general circulation models	
6.1.1.2 **Reanalysis**	This component concerns the software, processes and governance mechanisms that establish "a meteorological data assimilation project which aims to assimilate historical observational data spanning an extended period, using a single consistent assimilation (or "analysis") scheme throughout". (See Wikipedia article on meteorological reanalysis)	Optional
6.1.1.3 **Model ensembles**	This component refers to the software, processes and governance mechanisms used to aggregate data from: • A number of GCMs to produce products that portray a range of model forecasts. • A series of runs of the same model.	Optional

6.1.2 Generate derived data from climate observations

6.1.2.1	This component represents the software, processes and governance tools that handle a	Recommended

Spatial analysis	very wide variety of raster and vector spatial analysis techniques. Some examples are: • Generating grids that show the spatial distribution of observations of a phenomenon such as precipitation. • Generating grids that represent the distribution of the average maximum temperature for the month of May for climatological standard normals. • Generating grids that represent the distribution of the maximum temperature anomalies for May 2010 when compared to the climatological standard normal. • Selecting all meteorological stations located within a 10 km radius around a national administrative boundary.	
6.1.2.2 **Image analysis**	This component covers the software, processes and governance tools that handle a very wide range of image analysis techniques. Some examples are: • Processing remotely sensed satellite imagery to measure the relative solar reflectance of a satellite image, determine the cloud cover within a scene or generate a normalized difference vegetation index to measure vegetation greenness. • Processing ground-based radar imagery to detect rain and storm activity.	Recommended
6.1.2.3 **Time-series analysis**	This component concerns the software, processes and governance mechanisms that analyse time-series data using a very broad range of analysis techniques. Some examples are the analysis required to produce: • WMO standard products such as extremes, standard normals, World Weather Records and climate change indices. • A variety of derived observations data. For more information, see: • *Guide to Climatological Practices* (WMO-No. 100), Chapter 4 Characterizing climate from datasets, and Chapter 5 Statistical methods for analysing datasets	Recommended
6.1.2.4 **Teleconnection**	This component represents the software, processes and governance processes used to analyse, record and manage data representing teleconnections and major climate indices such	Optional

indices	as the El Niño-Southern Oscillation and the Southern Oscillation Index. According to an article on Wikipedia, "[t]eleconnection in atmospheric science refers to climate anomalies being related to each other at large distances (typically thousands of kilometers)".	

6.1.3 Data homogenization

6.1.3.1 **Data homogenization**	This component refers to the processes, software, governance and analysis of high-quality observations data and metadata used to develop 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. For more information, see: • *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186), WCDMP-53 • Trewin (2012) provides a good case study • ETCCDI website • The different homogenization packages available on the data-rescue website of the WMO Regional Association VI (RA VI) and CCl Task Team on Data Rescue (TT-DARE) • The European project – Advances in homogenisation methods of climate series: an integrated approach HOME (ACTION COST-ES0601, 2014)	Recommended

6.1.4 Other

6.1.4.1 **Other**	This component concerns as yet undefined processes, software and governance mechanisms used to analyse data.	Optional

7 Climate data presentation
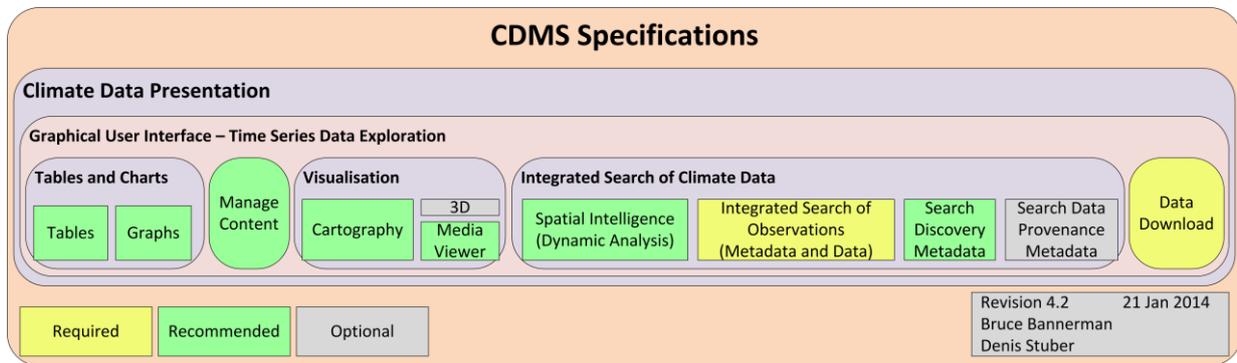
Figure 9. Climate data presentation

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 issues 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.	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 6 Services and 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: • Preparing texts, documents and data for effective web presentation. • Using technology such as content management systems to simplify web content presentation. • Implementing effective governance processes that review, validate and authorize content prior to being published.	Recommended
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 issues. It includes: • Spatial data preparation • Cartography • Simple point-and-click web 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	This component represents the technology, software, processes and governance processes	Recommended

Spatial intelligence	that support an effective and dynamic analysis of climate data within a web environment to facilitate understanding of climate matters and communicate issues relating 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 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
		Required

	– 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 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 key words. • Review discovery metadata records that adequately describe a dataset to enable searchers to determine whether it 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. For more information, see: • Section on climate metadata (4.3) • Subsection on dataset discovery metadata (4.3.2) • Discovery metadata catalogue component (8.2.1.1) • Linked data component (8.2.2.1)	Recommended
7.1.4.4 **Search data**	This component provides the functionality that allows an end-user to search the CDMS data provenance metadata catalogue to:	Optional

provenance metadata	• 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:
	• 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)
		94

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
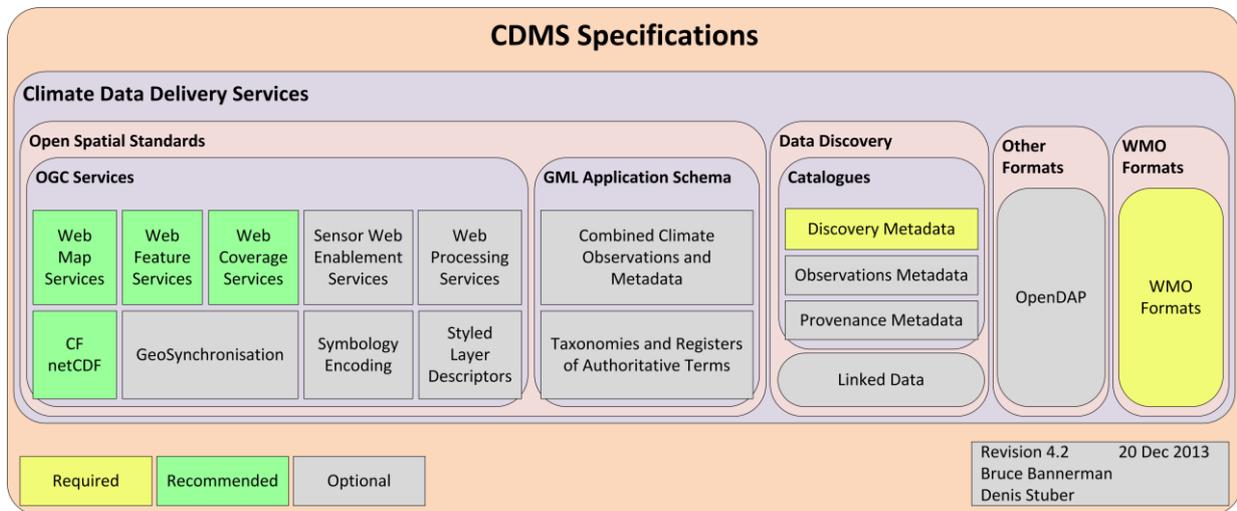
Figure 10. Climate data delivery services

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 Climate Challenge Integration Plugfest 2009 executive summary video, which describes the results of a global collaborative project demonstrating how climate data could be used via open spatial standards.

- The Open-source Geospatial Foundation overview of OGC standards (see OSGeo Live).

- The Geonovum wiki, which provides an overview of open spatial standards. This wiki is an initiative of the National Spatial Data Infrastructure executive committee in the Netherlands.

- The OGC Abstract Specifications, which provide a more 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 - The OGC Meteorology and Oceanography Domain Working Group paper on the use of WMS with time-dependent and elevation-dependent data	Recommended
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 GML (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	Recommended

	Service may be better for time-series observations. For more information, see: • OGC WFS documentation	
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 distribution of a wide variety of observational data and related metadata. There are a number of services that will become available when work stabilizes on these standards. 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[3]. For more information, see: • Wikipedia article on netCDF • OGC netCDF standards suite • The OGC CF-netCDF core and extensions primer	Recommended
8.1.1.6	This component concerns a series of technological tools that enable a data publisher to	Optional

---

[3] 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.

Geosynchronizati on	distribute a data product in an environment that supports managed change to the source data. In theory, end-users could subscribe to a data source and have changes to the data replicated in their copy of the data. 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 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. • To enable a future service-provider to offer CDMS-related services via "the cloud". For more information, see: • WPS Wikipedia entry • OGC WPS documentation	Optional

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. 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	Optional

8.1.2 Geography Markup Language application schema

Application schemas[4] (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.

Work is proceeding within WMO to harmonize existing TDCFs (FM 92 GRIB Edition 2 and FM 94 BUFR Edition 4) with WMO METCE (described in the METCE component (4.2.3.1)), with the intent to bind those existing data formats to explicit, well-understood semantics.

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

---

[4] Also known as logical data models

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. For an example, see the Combined climate observations and metadata component (8.1.2.1) below.

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.

While work has been conducted for several years, this task is still in its early stages as at December 2013. It is expected to take a number of years to complete.

For an overview of how a logical data model (and associated application schema) could be used with climate data, see Bannerman (2012), pp. 20–26.

For a more detailed description of what has been done to date, see Tandy (2013*a*), which provides an overview of the direction that WMO logical data model work is taking with regard to METCE.

8.1.2.1 **Combined climate observations and metadata**	This component represents the technology, software, processes and governance needed to support the transmission, consumption and processing of combined climate observations and associated metadata via a future climate observations application schema (or similar name) derived from: • The schema outlined in the Climate observations application schema component (4.2.3.2) • The schema outlined in the METCE component (4.2.3.1) As discussed in the section above entitled GML application schema (8.1.2), this work is currently embryonic. However, it is anticipated that it will become a key technological mechanism for exchanging climate observations and metadata in future years in support of data interoperability and platform independence. For more information, see: • Above section (8.1.2) • METCE component (4.2.3.1) • Climate observations application schema component (4.2.3.2)	Optional
8.1.2.2 **Taxonomies and registers of authoritative**	This component is related to the previous component. It represents the technology, software, processes and governance needed to develop an authoritative definition of the concepts and terms referenced in a logical data model such	Optional

terms	as a future climate observations application schema (or similar name), and to enable the publication of such terms.	
	Following the work of the Task Team on Aviation XML, WMO has established the WMO Codes Registry, which provides a web-based publication of terms from the *Manual on Codes* (WMO-No. 306).	
	The current coverage of terms is sparse – only covering the aviation-related terms required by the sponsoring activity – but there is commitment from WMO to populate the remaining code tables.	
	The WMO Codes Registry provides a well-defined programmatic application programming interface (API) alongside the web application. Where the need arises for publication of locally managed terms, it is recommended that the registry API be supported.	
	An open-source reference implementation of the registry software is available.	
	For more information, see: • Above section entitled GML application schema (8.1.2) • METCE component (4.2.3.1) • Climate observations application schema component (4.2.3.2) • Overviews of the WMO Codes Registry (Tandy, 2013*b*, 2013*c*)	

8.2   Data discovery

8.2.1 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 on-line services may be used to access each dataset. A discovery metadata catalogue allows an organization to participate within the WIS environment.	**Required**
8.2.1.2	This component refers to the technology and processes that create the observations metadata	Optional

Observations metadata catalogue	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.	
8.2.1.3 **Data provenance metadata catalogue**	This component refers to the technology and processes that create the data provenance metadata catalogue used to publish an organization's data provenance metadata records. It is anticipated that the climate database will be used to store and manage data provenance metadata. This component will serve as an IT catalogue service for data provenance metadata. More work is required to define this component.	Optional

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 http://lab.environment.data.gov.au/ 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: • A presentation on linked data (Berners-Lee, 2009) • Article on the ACORN-SAT linked climate dataset (Lefort et al., 2013)	Optional

8.3   Other formats

8.3.1 Open-source Project for a Network Data Access Protocol

8.3.1.1	This component represents technology suitable	Optional

OPeNDAP	for the distribution of a wide range of scientific data via the Data Access Protocol (DAP). DAP is used in the world's scientific community to allow Internet access to a range of scientific data. The protocol does not support the concept of spatial reference systems as defined in the OGC Abstract Specification on spatial referencing by coordinates. This makes it very difficult to reliably integrate data hosted via DAP with other spatial datasets, including those hosted via open spatial services. Therefore, it is considered more appropriate for cases in which researchers and scientists only want the data for numerical analysis. For more information, see: • OPeNDAP Wikipedia page • DAP 2.0 Specification	

8.3.2 WMO formats

8.3.2.1 **WMO formats**	This component represents technology suitable for the distribution of a wide range of climate data via traditional WMO formats. 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**

9 Core IT infrastructure
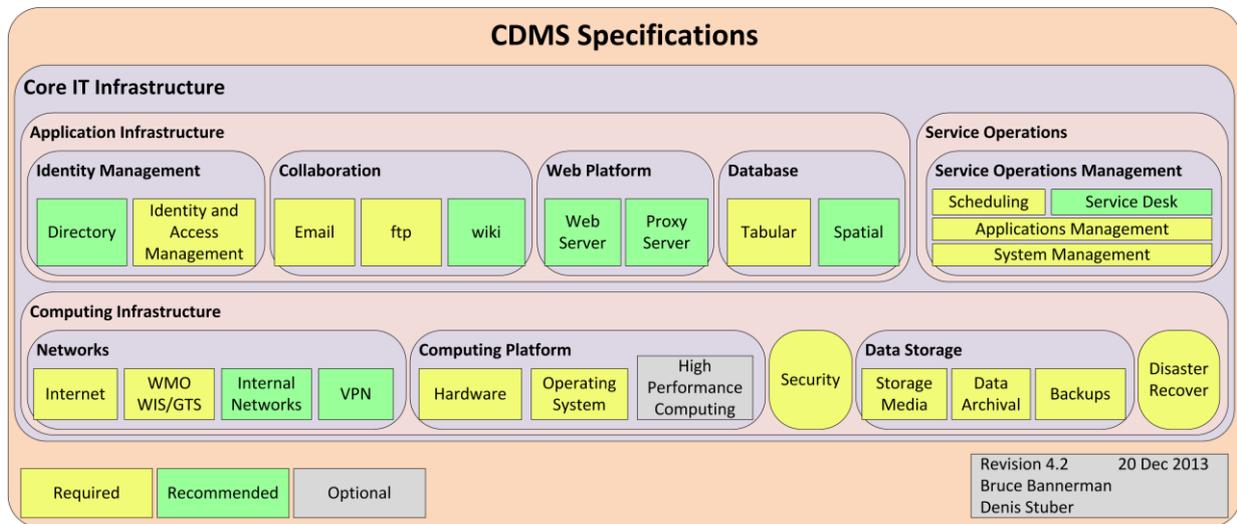
Figure 11. Core IT infrastructure

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.	**Required**
9.1.2.2	This component provides secure services to allow exchange of climate data via the use of the File Transfer Protocol (FTP).	**Required**

FTP		
9.1.2.3 **Wiki**	This component supports a collaborative web environment allowing any member of a team to easily edit content.	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 over a 24-hour basis. 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[5], 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.	Recommended
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. 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**

9.3   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 Global Telecommunication System. 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	Recommended

---

[5] See also the ITIL publication on service operations.

	Configuration Protocol.	
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

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.	**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.	Required

9.3.5 Disaster recovery

9.3.5.1 Disaster recovery	This component refers to the disaster recovery and business continuance 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.	Required

10 Considerations

There are many aspects that organizations are advised to consider when implementing a 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. 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 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.

It is suggested that the team:

- Have a clearly understood mandate to perform the work.
- 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 and project plan are developed.
- 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

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.

In addition, it would facilitate the planning process if team members could identify potential CDMS implementation costs during planning activities.

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 business?
    - How is that value measured and monitored?
- Business requirements
    - What business need is being met by current CDMS processes?
    - Which of the CDMS components described in this publication are being met by current approaches?
    - What CDMS governance processes exist?
- Data issues
    - What data holdings are being managed?
    - What is the format of the data?

- – Where are the data 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 by observations data?
- – What data licence 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 licences 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?
  - – 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?

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 need.

Service designers should have a clear 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 Specification 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.

10.4.2 Determine non-functional requirements

In addition to the functional requirements, a number of non-functional requirements (see Wikipedia article) relating to the operation of the CDMS will also need to be considered.

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.
- 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, due 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. 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 with climate data. This implies the need for a solution that uses open-source 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.
- 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, etc., or possibly 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?

- Can the organization afford to maintain the existing CDMS in addition to the new CDMS, both in terms of IT and staff skills?

- 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?

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?

10.6.4 Training

What training is required to ensure that staff can effectively deploy, manage, maintain and use the new CDMS?

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.

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.

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?

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

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.

11 Recommendations

11.1 Establish the WMO Climate Data Framework

Recommendation: Establish the WMO Climate Data Framework (CDF) to provide a strategic CDMS policy and governance structure that will underpin future consistent global climate services.

Note: The WMO CDF concept outlined below is intended to facilitate the availability of and access to consistent climate data across all geographical scales for WMO Members. The concept is expected to provide a major contribution to the broader High Quality Global Data Management Framework for Climate, which is currently under development within the WMO community.

In time, the CDF will drastically reduce the work required to undertake global and regional climate-related analysis by facilitating consistent approaches to policy, definition, management and use of climate data and related services.

This goal will, however, require data and service providers to commit to following the WMO CDF policies and approaches.

11.1.1 Why is the Climate Data Framework needed?

It is becoming increasingly important to obtain a consistent view of climate data to facilitate regular comprehensive analysis of global and regional climate-related issues.

This is currently very difficult to do, and analysts must waste considerable time massaging and transforming data into a form suitable for analysis. Some examples of issues faced are:
- Basic concepts:
    - What is the definition of a climatological day? Some data providers use 9 a.m. to 9 a.m. the following day, while others use 12 a.m. to 12 a.m. Some have used various definitions over the life of a station.
    - How is daylight saving treated? Some continue to use the normal time zone while others switch to daylight saving time.
    - What period for climatological standard normals is used for regular analytical work? Some data providers maintain data according to the WMO standard normal of 1961–1990, others use differing standard normals.
- Consistent and objective quality assurance measures:
    - Different data providers use different approaches to quality assurance, with varying and inconsistent measures of quality. Most of these measures are subjective, making it considerably difficult to conduct an objective analysis.
    - There is no standardized way to measure the uncertainty inherent in climate data.
- Consistent view of climate metadata:
    - There is no clear and consistent understanding of what comprises climate metadata. For example, different countries maintain differing observations metadata, in part due to WMO documentation that cannot be easily implemented in an IT system.

- There is widespread confusion when the concept of metadata is discussed; it makes some think of discovery (or WIS) metadata and others of observations (or WIGOS) metadata.
- Inconsistent approaches to common data problems:
    - Many different approaches are used to solve a wide range of data problems, including:
        - How to account for missing data when creating aggregated derived datasets.
        - How to treat semantic differences in data when data observation techniques or coding standards have changed with time.
        - How to treat datasets in which the data are composed of observations of varying levels of quality.
- Taxonomy:
    - There is no authoritative taxonomy that can be used to consistently define a wide range of climate-related terms. This leads to a number of issues when attempting to integrate data from different providers. The WMO Codes Registry is starting to address this issue.
- Consistent datasets:
    - There are a number of data consistency issues, for example:
        - Each data provider maintains data in disparate sources, using different data models with varying data definitions.
        - In addition, there is no uniform format and structure for delivering data. Even when data appear to be the same, there are often semantic differences between the datasets.
        - With the exception of WMO standard products, there is no real concept of standardized climate data products, with the same definition and structure used by all data providers.
    - Many countries are currently unable to provide even basic products, including mandatory WMO reports such as CLIMAT and World Weather Records. In this publication, these products are classified as required for a CDMS. There may be other products that are desirable as well.
- Consistent services:
    - Similarly, there are currently no consistent international practices for the delivery of climate data and products. It is argued that an open spatial standards approach would facilitate automated integration and processing of data from different providers.

A CDF that data providers and participants commit to following could help address the majority of the issues discussed above. A CDF could eventually lead to:

- Consistent climate data products and services based on open standards that can be readily combined and analysed for a range of climate applications with minimal wasted time and effort.
- Consistent climate software that can be easily integrated and adopted by NMHSs and other interested participants.
- Less ambiguity in climate data issues through having clear and consistent policies that CDF participants can adopt.

There will be a cost for NMHSs in complying with a CDF, particularly with regard to implementing software that meets CDF requirements and reviewing and adjusting data

holdings and practices. However, as all software systems require maintenance and replacement at some stage, this is not seen as a significant barrier to the adoption of the CDF.

11.1.2 What is the WMO Climate Data Framework?

The concept of the WMO CDF has been derived from that of a spatial data infrastructure (SDI) (see Wikipedia article). Some examples of SDIs are the Infrastructure for Spatial Information in Europe, GEOSS, the Global Spatial Data Infrastructure and the United Nations Spatial Data Infrastructure.

Borrowing from the Global Spatial Data Infrastructure (GSDI) definition of an SDI (see GSDI Cookbook wiki), the WMO Climate Data Framework is defined as the base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to consistent climate data. The CDF provides a basis for climate data discovery, evaluation and application for users and providers within the WMO community, the commercial sector, the non-profit sector, academia and the global community in general.

WMO is well placed to sponsor and nurture the development of a CDF in that WMO:

- Has a broad membership that covers the majority of climate data providers.
- Has existing agreements in place with data providers and well-established governance processes for collaboratively facilitating such ideas.
- Is the global authority for climate data and issues through the Commission for Climatology and the Global Framework for Climate Services.
- Has existing collaborative agreements in place with key stakeholders such as ISO, OGC and GEO.
- Has effective processes in place for collaboration between WMO member organizations through a variety of expert teams.
- Has implemented the beginnings of an overarching SDI in WIS. Activities initiated under the CDF will eventually enhance WIS with climate domain specific requirements.

It is anticipated that the CDF could be developed within the existing Commission for Climatology OPACE structure.

To ensure widespread acceptance of the CDF, it is suggested that WMO welcome collaboration with other interested external organizations where appropriate.

11.1.3 What work could facilitate the emergence of the Climate Data Framework?

The following set of tasks is suggested to facilitate the emergence of the CDF:

- Develop a clear understanding of the CDF and its future development path, goals and timelines.
- Establish the CDF governance processes for creating, reviewing, accepting, maintaining and committing to CDF policies.
- Develop consistent policies as per Chapter 3 of this publication, with additional policies as deemed appropriate by the CDF implementation team.
- Use this CDMS Specification as a starting point and facilitate further development of CDMS specifications.

- List the current data inconsistency issues causing problems with data integration and develop an approach to resolve the identified issues.
- Develop and define objective measures of climate data quality that can be implemented in IT systems, as outlined in Chapter 5 of this publication. Of particular note is the requirement to establish objective classification schemes for:
  – Sustained performance (component 5.4.1.2)
  – Multilayer quality flags (component 5.4.1.3)
  – Climate observation quality (component 5.4.1.4)
  – Derived-data quality assessment (component 5.4.2.1)
- Define and design consistent data product specifications for standard climate products in accordance with ISO 19131 *Geographic information – Data product specifications* (see ISO/TC 211 Advisory Group on Outreach *Standards Guide: ISO/TC 211 Geographic Information/Geomatics*). This will become the foundation for future consistent climate data services based on open standards.
- Continue to liaise with other commissions and working groups.
- Develop a clear, consistent definition of climate metadata including observations, discovery and data provenance metadata. Ensure that this definition is sufficient to enable the implementation of appropriate software that will support the required functionality. In doing this work:
  – Further develop and arrive at a common understanding of the concept of observations metadata with the appropriate WIGOS working groups.
  – Further develop and arrive at a common understanding of the concept of discovery metadata with the relevant WIS and Commission for Basic Systems working groups.
  – Collaborate with the W3C community to further develop the concept of data provenance metadata as required.
  – Liaise with the OGC and ISO to ensure that climate requirements are met within open spatial standards and services.
- Coordinate with WIS groups to ensure that CDF work integrates with and enhances WIS and vice versa.
- Work with the Inter-Programme Expert Team on Metadata and Data Representation Development (IPET-MDRD) and related teams to develop the WMO logical data model and ensure that it caters for climate data and climate metadata requirements.
- Also develop appropriate climate-related taxonomies for the WMO logical data model and related future services and publish them in the WMO Codes Registry. The naming of observation variables, for example, is especially pressing.

11.1.4 Relationships between the Climate Data Framework and other spatial data infrastructures

It is anticipated that in time, the Climate Data Framework will become the authoritative definition of climate data and related technologies, policies and institutional arrangements that facilitate effective analysis and use of consistent climate data.

Once this has occurred, the CDF could be adopted by other SDIs as the authoritative definition of climate data requirements. Relevant existing SDIs that could benefit from CDF work (and vice versa) are:

- WIS

- Global Spatial Data Infrastructure
- United Nations Spatial Data Infrastructure
- Infrastructure for Spatial Information in Europe
- National spatial data infrastructures, such as the Australian Spatial Data Infrastructure or the US National Spatial Data Infrastructure

11.2 Continue developing climate data management system specifications

Recommendation: Continue developing the CDMS Specification to provide an authoritative definition of the functionality expected within a CDMS.

11.2.1 Why is additional specification work needed?

This version of the CDMS Specification is intended to provide NMHSs with a strategic vision of the functionality required within a CDMS.

Sufficient detail has been included to help NMHS technical experts in both climate and IT domains to evaluate the functionality delivered by their existing CDMS and start planning activities to upgrade their CDMS as needed.

Additional work is required to ensure that a future version of this CDMS Specification can become the stand-alone authoritative definition of CDMS requirements. Once this has been achieved, regular maintenance will be necessary to ensure that the CDMS Specification remains up to date.

The following are examples of areas requiring additional work:
- Expanding and refining each CDMS component to ensure that:
    - It contains enough information to allow the component's functionality to be properly implemented in an IT system.
    - The information contained in the CDMS Specification is the best available.
    - The Specification can be read as a stand-alone document – readers are not directed to other references in order to find relevant information.
- Ensuring that each component is clearly and unambiguously defined. This will allow participating countries to determine whether they have or have not complied with what is required, recommended or optional.
- Making sure that each component is explicit, for example by avoiding the use of "etc" or "and so forth" to extend a list.
- Ensuring consistent styling of component statements. Currently, some components provide explicit information while others comprise questions to get readers to think about relevant issues.
- Ensuring that observations metadata are extended to include the requirements of remote observation platforms, such as satellites and airborne or marine sensors.
- Developing relevant use cases (see Wikipedia article) and processing swim lanes (see Wikipedia article), together with other suitable documentation that define how the functionality is expected to be used.

11.3 Establish and maintain a register of known climate data management systems

Recommendation: Establish and maintain a CDMS register that contains the results of an evaluation of how well each CDMS meets the requirements of this CDMS Specification.

11.3.1 Why is a register needed?

This register is required in order to support NMHSs that need to upgrade their CDMS.

Prior to this CDMS Specification, there was no consistent definition of what functionality is expected within a CDMS. Consequently, current systems that are described as a CDMS contain varying functionalities. Very few, if any, will meet the requirements of this Specification.

Rather than have a number of NMHSs try to find suitably qualified and experienced experts and independently evaluate a finite set of CDMSs, it would be more beneficial to conduct a combined evaluation and make it available for all NMHSs to use as a reference source.

Considering that all CDMSs may be undergoing change to improve their functionality, it is further recommended that, once the register is established, an annual review be conducted to ensure that the register retains its relevance.

11.4 Sponsor and support open-source approaches to climate data management system development

Recommendation: Noting the requirements and resource limitations of developing countries, sponsor and support open-source approaches to CDMS development and establish communities of practice.

11.4.1 Why support an open-source approach to climate data management system development?

An open-source approach to CDMS development is desirable in order to:
- Spread the cost, risk and burden of developing and maintaining CDMSs among a number of organizations.
- Ensure that CDMS investment is not locked up in black-box, closed-source solutions and that the intellectual property and source code of developed software can be reused as required, with little or no restriction.
- Ensure that CDMS software is open to review and scrutiny so as to foster trust in the work being done to process climate data.
- Facilitate the transfer of knowledge to those interested in learning more about CDMSs so that they may assist with software development and maintenance.
- Develop self-sustaining open-source communities to eventually support CDMS software components over the long term.

11.4.2 Open-source versus free

A common misconception is that open-source software is developed and maintained by volunteers and is therefore free of costs.

This is not necessarily the case. The term "free" that is often used in conjunction with "open-source software" refers mainly to the idea that the use of the software is free from any constraints. Of particular relevance is the freedom to access and modify the source code and redistribute it with minimal restrictions, often at no cost.

There is, however, always a cost associated with the development and maintenance of open-source software. Increasingly, this cost is being met by organizations (commercial, government, research or other) and spread across a number of parties.

For further information, see Wikipedia articles on:
- The ambiguities of the term "free"
- Free and open-source software
- Free software

11.4.3 Open-source projects

This CDMS Specification is intended to support a wide variety of approaches to CDMS development, from closed-source, proprietary solutions to open-source methods, or a combination of the two. In fact, the functionality of many of the CDMS components presented in this publication could be delivered by existing, well-established closed- or open-source software applications.

Consideration should be given to the findings of the UK parliamentary inquiry and in particular to the following:

> 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). (UK Parliament Science and Technology Committee, 2011)

This implies that the algorithms and software that process climate data should be available for scrutiny. This can be most effectively achieved by adopting open-source software practices with the CDMS.

Taking a pragmatic approach, this does not imply that all CDMS components should be open-source. Effective arguments could be made for using closed-source components that underpin the CDMS, such as closed-source core IT infrastructure components that may better suit NMHS requirements and the existing infrastructure.

It should also be noted that adopting an open-source approach does not stop or restrict commercial involvement in a CDMS. To the contrary, many commercial organizations, including very large multinationals, routinely sponsor and promote the use of open-source software. Open-source projects can be a very effective way to spread the costs and risks of software development and maintenance among a number of parties. Many commercial organizations are also finding it profitable to offer commercial support for such software.

CDMSs can be very expensive to develop and maintain. They also require a constant pool of very talented and experienced people with skills in a number of areas.

It is expected that NMHSs may need to invest in purchasing or developing new CDMS functionality. It is therefore recommended that resources from WMO, NMHSs, commercial organizations and other interested parties be pooled to share the burden of developing and

maintaining CDMSs. Such resources could include funding, IT, climate data management or scientists' time and expertise.

This effort will be most effective if undertaken and coordinated via appropriately structured open-source projects in which all communications, planning, development and governance processes occur within an open and public forum. This will foster open collaboration by participating organizations or individuals and encourage the transfer of knowledge between participants. It is further anticipated that such projects will generate interest from the wider open-source community who may be encouraged to contribute their knowledge and expertise.

It is noted that there are currently several CDMSs in the early stages of open-source development, for example the Climate Data for the Environment (CLiDE) database and the Meteorology, Climatology and Hydrology (MCH) system. Both products:

- Are in their early stages of development, with limited functionality available.
- Are focused on delivering the needs of their funding bodies, with limited external review.
- Have not yet begun to try to establish an open-source community to support the CDMS development.
- Have very limited development support and little funding.
- Are at high risk due to:
  - Key-person dependency – this would cause problems if key developers accepted alternative employment, became ill or left for other reasons.
  - Limited funding – should funding cease to be provided to either CDMS in its current form, its development teams would likely be disbanded and move on to other activities. This would make it very difficult to restart the CDMS development.

This situation has emerged due to the short-term, project-based approach to funding that the CDMSs have been developed under. Even though the software has been released under an open-source licence, minimal effort has been made to establish open-source communities to ensure widespread understanding of the design and structure of the software components and to support the projects over the long term.

This situation cannot be allowed to continue. A commitment is needed to sponsor and fund relevant open-source CDMSs for a number of years so that open-source communities can be established with a view to becoming self-sustaining. This work will need to be coordinated, possibly by the new ET-CDMS or a similar mechanism.

11.4.4 First steps

Developing countries have the same requirement for basic CDMS functionality as developed countries. Considering the many issues the former are currently facing to fund, establish and maintain effective CDMSs, establishing open-source CDMSs that support the needs of such countries is considered a good place to start.

As there is considerable potential for the development of open-source CDMSs that could be scaled to meet the needs of both developing and developed countries, such a factor should be included in the design of the CDMS to enable its future extension.

It is too early to identify a single open-source CDMS as a superior product. Therefore, it is recommended that the creation of several open-source CDMSs be supported to see which can

attract an open-source community to sponsor and fund its development and maintenance over the long term. Market realities will help with the rationalization of open-source CDMSs and components in the future.

Both NMHSs and commercial organizations should be able to apply for funding. The competition for funding will encourage appropriate CDMS functionality (as per this CDMS Specification) and openness of development practices, assuming that proper conditions are attached to the funding.

If implemented effectively, an open-source approach would lead to the establishment and growth of open-source communities, enabling NMHSs, commercial organizations and other interested parties to pool resources and collaborate on their development.

This approach could fast-track CDMS development activities and eventually result in a community that could support the CDMS over the long term.

11.5 Fast-track the development of the WMO logical data model

Recommendation: Fast-track the development of the WMO logical data model to support the exchange of combined climate observations and climate metadata.

11.5.1 Why is the WMO logical data model needed?

The WMO logical data model is described in subsection 4.2.3 of this publication. Further references are made in subsection 8.1.2, relating to the development of GML application schemas for climate observations.

The WMO logical data model and its future support for the exchange of combined climate observations data and climate metadata (observations, discovery and data provenance) is considered a key WMO development that will:

- Help establish authoritative taxonomies of climate concepts.
- Facilitate effective exchange of climate data via an independent logical data model and associated XML representation (GML application schema).
- Facilitate interoperability between climate applications and components by providing a consistent CDMS-independent view of climate observations data.
- Underpin future climate services.

Work on the WMO logical data model will also require consistent data policies, definitions and other elements described in recommendation 11.1.

WMO logical data model development is taking place within IPET-MDRD (see IPET-MDRD terms of reference) and the IPET-MDRD Task Team on Data Modelling and Representation (TT-DMR) (see TT-DMR terms of reference).

Other organizations outside of WMO are also interested in this concept. An example is the preliminary work conducted by the OGC Meteorology and Oceanography Domain Working Group, who will be interested in collaborating with WMO in the development of the WMO logical data model.

11.6   Define a globally unique WMO identifier for stations

Recommendation: Fast-track the development of globally unique WMO identifiers.

11.6.1 Why is a globally unique WMO identifier needed?

The current WMO station identifier system is inadequate and results in confusing and inappropriate numbering practices. For example, some NMHSs are running out of station identifiers and reusing the identifiers previously allocated to other stations.

This is likely to create problems with global time-series analysis, particularly when analysts are not aware that a station identifier has been reused at another location.

A way needs to be found to fast-track this key WMO initiative, as it is a vital strategic component that will underpin future climate services.

References

WMO sources

(a) References

Bannerman, B., 2012: Stations metadata and WMO Core Profile – A way forward. Discussion paper, ET-CDMS, World Meteorological Organization. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/etcdms-metadata-discussionpaper.pdf.

Stuber, D., 2012: Climatological needs for minimum station metadata in the frame of the WMO's publication 9 volume A. Discussion paper, ET-CDMS, World Meteorological Organization. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/Station_Metadata_and_VolumeA.pdf.

Tandy, J., 2010: WMO Core Metadata Profile version 1.2 – Guidelines on the use of metadata for WIS. Draft, Open Programme Area Group on Information Systems and Services, World Meteorological Organization. http://wis.wmo.int/2010/metadata/version_1-2/WMO%20Core%20Metadata%20Profile%20v1-2%20Guidance%20Documentation%20v0.1%20(DRAFT).pdf.

———, 2013*a*: Development of a GML application schema for data exchange supporting meteorological services for international air navigation. PowerPoint presentation, Task Team on Aviation XML, World Meteorological Organization. http://old.ecmwf.int/newsevents/meetings/workshops/2013/GIS-OGC_standards/Presentations/pdfs/Tandy.pdf.

———, 2013*b*: WMO Codes Registry. PowerPoint presentation, Task Team on Aviation XML, meeting 3, 7–9 October, Montreal. World Meteorological Organization. http://wis.wmo.int/file=521.

———, 2013*c*: WMO Codes Registry: web-based publication of the *Manual on Codes*. PowerPoint presentation, World Meteorological Organization. http://old.ecmwf.int/newsevents/meetings/workshops/2013/MOS14/Presentations/Tandy.pdf.

World Meteorological Organization, 1967: *A Note on Climatological Normals: Report of a Working Group of the Commission for Climatology*. Technical Note No. 84. Geneva. Not available online.

———, 1975: *International Cloud Atlas*, Volume I (WMO-No. 407). Geneva. http://library.wmo.int/pmb_ged/wmo_407_en-v1.pdf.

———, 1986: *Guidelines on the Quality Control of Surface Climatological Data* (Abbott, P.F.). WCP-85 (WMO/TD-No. 111). Geneva. http://library.wmo.int/opac/index.php?lvl=notice_display&id=11700.

———, 1989: *Calculation of Monthly and Annual 30-Year Standard Normals*. WCDP-10 (WMO/TD-No. 341). Geneva. http://www.inmet.gov.br/html/clima/OMM_WCDP_N10.pdf.

———, 1993: *Guide on the Global Data-processing System* (WMO-No. 305). Geneva. http://library.wmo.int/pmb_ged/wmo_305_en.pdf.

———, 1996: *1961–1990 Global Climate Normals (CLINO)* (WMO-No. 847). Geneva. http://library.wmo.int/opac/index.php?lvl=notice_display&id=7785.

———, 2003*a*: *Guidelines on Climate Metadata and Homogenization* (Aguilar, E., I. Auer, M. Brunet, T.C. Peterson and J. Wieringa, edited by P. Llansó). WCDMP-53 (WMO/TD-No. 1186). Geneva. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-53.pdf.

———, 2003*b*: *Guidelines on Climate Observation Networks and Systems* (Plummer, N., T. Allsopp and J.A. Lopez, edited by P. Llansó). WCDMP-52 (WMO/TD-No. 1185). Geneva. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-52_000.pdf.

———, 2004: *Guidelines on Climate Data Rescue* (Tan, L.S., S. Burton, R. Crouthamel, A. van Engelen, R. Hutchinson, L. Nicodemus, T.C. Peterson and F. Rahimzadeh, edited by P. Llansó and H. Kontongomde). WCDMP-55 (WMO/TD-No. 1210). Geneva. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-55.pdf.

———, 2007*a*: *Guidelines on Climate Data Management* (Plummer, N., W. Lipa, S. Palmer, G. Prank, J. Shortridge and D. Stuber, edited by O. Baddour and H. Kontongomde). WCDMP-60 (WMO/TD-No. 1376). Geneva. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMPNo60.pdf.

———, 2007*b*: *The Role of Climatological Normals in a Changing Climate* (Trewin, B., edited by O. Baddour and H. Kontongomde). WCDMP-61 (WMO/TD-No.1377). Geneva. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMPNo61.pdf.

———, 2008: *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8). Geneva. http://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf.

———, 2009*a*: *Handbook on CLIMAT and CLIMAT TEMP Reporting* (WMO/TD-No. 1188). Geneva. http://www.wmo.int/pages/prog/www/OSY/Publications/TD1188/HandbookCLIMAT-CLIMATTEMP_en.pdf.

———, 2009*b*: *Practical Help for Compiling CLIMAT Reports.* GCOS-127 (WMO/TD-No. 1477). Geneva. http://www.wmo.int/pages/prog/gcos/Publications/GCOS-127_EN.pdf.

———, 2010*a*: *Guide to the Global Observing System* (WMO-No. 488). Geneva. http://library.wmo.int/pmb_ged/wmo_488-2010_en.pdf.

———, 2010*b*: *Manual on the Global Data-processing and Forecasting System*, Volume I (WMO-No. 485). Geneva. http://www.wmo.int/pages/prog/www/DPFS/documents/485_Vol_I_en.pdf.

———, 2010*c*: *Manual on the Global Observing System*, Volume I (WMO-No. 544). Geneva. https://docs.google.com/file/d/0BwdvoC9AeWjUV2dIQmlSUkpOYm8/edit.

———, 2011*a*: *Guide to Climatological Practices* (WMO-No. 100). Geneva. https://docs.google.com/file/d/0BwdvoC9AeWjUM3RhLUszWDdaTUU/edit.

———, 2011*b*: *Manual on Codes*, International Codes, Volume I.1, Part A (WMO-No. 306). Geneva. http://library.wmo.int/pmb_ged/wmo_306-v1_1-2012_en.pdf.

———, 2011*c*: *Manual on Codes*, International Codes, Volume I.2, Parts B and C (WMO-No. 306). Geneva. http://library.wmo.int/pmb_ged/wmo_306-v1_2-2012_en.pdf.

———, 2011*d*: *Manual on Codes*, International Codes, Volume I.2, Part C, section d. (WMO-No. 306). Geneva:

- http://www.wmo.int/pages/prog/www/WMOCodes/BC_Regulations/BC30-CLIMAT.pdf
- http://www.wmo.int/pages/prog/www/WMOCodes/BC_Regulations/BC32-CLIMATSHIP.pdf
- http://www.wmo.int/pages/prog/www/WMOCodes/TemplateExamples.html#Regulations

———, 2011*e*: *Manual on the Global Telecommunication System*, Volume I (WMO-No. 386). Geneva. https://docs.google.com/file/d/0BwdvoC9AeWjURy1HUDlweXhKWXc/edit.

———, 2012*a*: Final report of the first session of the Commission for Instruments and Methods of Observation Expert Team on Standardization, 26–29 November, Geneva. http://www.wmo.int/pages/prog/www/IMOP/reports/2012/Final-Report_ET-Stand-1.pdf.

———, 2012*b*: *Guidelines on Submission of the World Weather Records Tenth Series (2001–2010)*. WCDMP-77. Geneva. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP_77.pdf.

———, 2012*c*: *Manual on Marine Meteorological Services*, Volume I (WMO-No. 558). Geneva. http://library.wmo.int/pmb_ged/wmo_558_en-v1.pdf.

———, 2012*d*: *Manual on the Implementation of Education and Training Standards in Meteorology and Hydrology*, Volume I (WMO-No. 1083). Geneva. http://library.wmo.int/pmb_ged/wmo_1083_en.pdf.

———, 2012*e*: *Manual on the WMO Information System* (WMO-No. 1060). Geneva. https://docs.google.com/a/wmo.int/file/d/0BwdvoC9AeWjUVWRoR1hPZjJOMDA/edit.

———, 2013*a*: *WMO Core Metadata Profile version 1.3,* Specification, Part 1. Open Programme Area Group on Information Systems and Services. Geneva. http://wis.wmo.int/2012/metadata/WMO_Core_Metadata_Profile_v1.3_Specification_Part_1_v1.0FINALcorrected.pdf.

———, 2013*b*: *WMO Core Metadata Profile version 1.3*, Specification, Part 2. Open Programme Area Group on Information Systems and Services. Geneva. http://wis.wmo.int/2012/metadata/WMO_Core_Metadata_Profile_v1.3_Specification_Part_2_v1.0FINAL.pdf.

———, 2013*c*: *Resolutions of Congress and the Executive Council* (WMO-No. 508). Geneva. https://docs.google.com/file/d/0BwdvoC9AeWjUYWJoUS1ienlOTEk/edit.

———: *Technical Regulations* (WMO-No. 49). Geneva:

- Volume I (2011) http://library.wmo.int/pmb_ged/wmo_49-v1-2012_en.pdf.

- Volume II (2010) http://library.wmo.int/pmb_ged/wmo_49-v2-2011_en.pdf.

- Volume III (2006) http://library.wmo.int/pmb_ged/wmo_49-v3-2006_en.pdf.

- Volume IV (2011) http://library.wmo.int/pmb_ged/wmo_49-v4_en.pdf.

———: *WIGOS – The WMO Integrated Global Observing System*. Geneva. Flyer, http://www.wmo.int/pages/prog/www/wigos/documents/Principal_Docs/WIGOS_flyer_en.pdf.

Wright, W., 2012*a*: Discussion paper on the calculation of the standard climate normals: a proposal for a dual system. Presented to the Management Group of the Commission for Climatology, World Meteorological Organization. https://www.wmo.int/pages/prog/wcp/wcdmp/documents/Rev_discussion_paper_May2012.pdf.

———, 2012*b*: Statement of guidance for climate – An analysis of current and emerging capacity gaps in surface and upper air observations to support climate activities. Expert Team for Evolution of the Global Observing System, World Meteorological Organization. http://www.wmo.int/pages/prog/www/OSY/SOG/SoG-Climate-CCl.doc.

(b) Web pages

AvXML-1.0 data model, http://wis.wmo.int/AvXML/AvXML-1.0/index.htm.

GCOS website, 2013, http://www.wmo.int/pages/prog/gcos/index.php?name=AboutGCOS.

GCOS ECVs, 2013, http://www.wmo.int/pages/prog/gcos/index.php?name=EssentialClimateVariables.

Terms of reference for IPET-MDRD, 2013, http://www.wmo.int/pages/prog/www/CBS/Lists_WorkGroups/CBS/opag%20iss/ipet-mdrd/tors.

Terms of reference for IPET-MDRD-TT-MDR, 2013, http://www.wmo.int/pages/prog/www/CBS/Lists_WorkGroups/CBS/opag%20iss/ipet-mdrd-tt-dmr/tors.

WMO Codes Registry, 2013, http://codes.wmo.int/.

WMO international codes, http://www.wmo.int/pages/prog/www/WMOCodes.html.

World Weather Records, http://www.wmo.int/pages/prog/wcp/wcdmp/GCDS_2.php.

Other sources

(a) References

Berners-Lee, T., 2009: *The next Web of open: linked data*. Video, http://blog.ted.com/2009/03/13/tim_berners_lee_web/.

ISO, 2005: *Geographic information – Rules for application schema*, ISO 19109:2005. Geneva. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=39891.

———, 2007: *Geographic information – Geography Markup Language (GML)*, ISO 19136:2007. Geneva. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32554.

———, 2009*a*: *Geographic information – Spatial referencing by coordinates, Part 2: Extension for parametric values*, ISO 19111-2:2009. Geneva. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=44075.

———, 2009*b*: *Standards Guide: ISO/TC 211 Geographic Information/Geomatics.* ISO/Technical Committee 211 Advisory Group on Outreach. http://www.isotc211.org/Outreach/ISO_TC_211_Standards_Guide.pdf.

———, 2011: *Geographic information – Observations and measurements*, ISO 19156:2011. Geneva. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32574.

———: *Geographic information – Data product specifications*, ISO 19131. Geneva. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=36760.

———: *Geographic information – Metadata,* ISO 19115. Geneva. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020.

Kern-Hansen, C., 2009: Summary of findings from workshop session. Seventh ECSN data management workshop at the Danish Meteorological Institute, 4–6 November, Copenhagen. http://www.dmi.dk/fileadmin/Rapporter/DKC/DKC09-10_chap03.pdf.

Lefort L., A. Haller, K. Taylor and A. Woolf, 2013: The ACORN-SAT linked climate dataset. *Semantic Web Journal*, http://www.semantic-web-journal.net/system/files/swj457.pdf.

OGC, 2010: OGC Abstract Specification: Spatial referencing by coordinates (Lott, R., ed.). http://portal.opengeospatial.org/files/?artifact_id=39049.

———, 2011*a*: CF-netCDF core and extensions primer (Domenico, B., ed.). http://portal.opengeospatial.org/files/?artifact_id=43733.

———, 2011*b*: OWS 7 engineering report – Geosynchronization service (Vretanos, P.A., ed.). http://portal.opengeospatial.org/files/?artifact_id=39476.

———, 2013*a*: OGC Abstract Specification: Geographic information – Observations and measurements (Cox, S., ed.). http://portal.opengeospatial.org/files/?artifact_id=41579.

———, 2013*b*: OGC best practice for using Web Map Services (WMS) with time-dependent or elevation-dependent data (Voidrot-Martinez, M.-F., C. Little, J. Seib, R. Ladner, A. Custer, J. de La Beaujardière and S. Siemen, eds.). OGC Meteorology and Oceanography Domain Working Group. http://external.opengeospatial.org/twiki_public/pub/MetOceanDWG/MetOceanWMSBPOnGoingDrafts/12-111r1_Best_Practices_for_WMS_with_Time_or_Elevation_dependent_data.pdf.

———, 2013*c*: *OGC Climate Challenge Integration Plugfest (CCIP) 2009*. Video, http://portal.opengeospatial.org/files/?artifact_id=36463.

Trewin, B., 2012: *Techniques Involved in Developing the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) Dataset*. Bureau of Meteorology, Melbourne, Australia. http://cawcr.gov.au/publications/technicalreports/CTR_049.pdf.

UK Parliament Science and Technology Committee, 2011: *The Reviews into the University of East Anglia's Climatic Research Unit's E-mails*. London. http://www.publications.parliament.uk/pa/cm201011/cmselect/cmsctech/444/44404.htm.

W3C, 2013*a*: PROV-DM – The PROV data model (Moreau, L., and P. Missier, eds.). http://www.w3.org/TR/prov-dm/.

- The PROV definition of provenance, http://www.w3.org/TR/2013/REC-prov-dm-20130430/#dfn-provenance.

———, 2013*b*: PROV-overview – An overview of the PROV family of documents (Groth, P., and L. Moreau, eds.). http://www.w3.org/TR/prov-overview/.

(b) Web pages

ACTION COST-ES0601, 2014, http://www.homogenisation.org/v_02_15/.

Creative Commons website, http://creativecommons.org/licenses/.

Creative Commons ShareAlike licence, http://creativecommons.org/licenses/by-sa/3.0/deed.en_US.

Data-rescue site of RA VI and CCI TT-DARE, 2014, http://www.meteobal.com/climatol/DARE/.

ETCCDI, 2013, http://etccdi.pacificclimate.org/indices.shtml.

Geonovum wiki, 2012, http://geostandards.geonovum.nl/index.php/Main_Page.

GEOSS, http://www.earthobservations.org/geoss.shtml.

GSDI Cookbook wiki, http://www.gsdidocs.org/GSDIWiki/index.php/Chapter_1.

ITIL website, http://www.itil-officialsite.com/.

ITIL lifecycle publication suite – books, http://www.best-management-practice.com/officialsite.asp?FO=1253138&ProductID=9780113313235&Action=Book&trackid=004037.

National Aeronautics and Space Administration, United States, DAP 2.0 Specification, http://earthdata.nasa.gov/our-community/esdswg/standards-process-spg/rfc/esds-rfc-004-dap-20.

OGC, Geography Markup Language (GML), http://www.opengeospatial.org/standards/gml.

———, Abstract Specifications, http://www.opengeospatial.org/standards/as.

———, Web Map Service (WMS), http://www.opengeospatial.org/standards/wms.

———, Web Feature *S*ervice (WFS), http://www.opengeospatial.org/standards/wfs.

———, Web Coverage Service (WCS), http://www.opengeospatial.org/standards/wcs.

———, Observations and measurements, http://www.opengeospatial.org/standards/om.

———, Sensor Observation Service, http://www.opengeospatial.org/standards/sos.

———, NetCDF standards suite, http://www.opengeospatial.org/standards/netcdf.

———, Geosynchronization services, http://www.opengeospatial.org/projects/groups/geosyncswg.

———, Web Processing Service (WPS), http://www.opengeospatial.org/standards/wps.

———, Symbology Encoding, http://www.opengeospatial.org/standards/se.

———, Styled Layer Descriptor (SLD), http://www.opengeospatial.org/standards/sld.

Old Weather, 2013, http://www.oldweather.org/.

OpenStreetMap, http://www.openstreetmap.org/.

OSGeo Live, http://live.osgeo.org/en/standards/standards.html.

Veilleurs du temps, http://www.veilleurs-du-temps.fr/pages/accueil.php.

Weather Observations Website (WOW), http://wow.metoffice.gov.uk/support/general.

Wikipedia, Backup, http://en.wikipedia.org/wiki/Backup.

———, Big data, http://en.wikipedia.org/wiki/Big_data.

———, BUFR, http://en.wikipedia.org/wiki/BUFR.

———, Climate, http://en.wikipedia.org/wiki/Climate.

———, Crowdsourcing, http://en.wikipedia.org/wiki/Crowdsourcing.

———, Enterprise architecture, http://en.wikipedia.org/wiki/Enterprise_architecture.

———, Epidemiology, http://en.wikipedia.org/wiki/Epidemiology.

———, Free and open-source software, http://en.wikipedia.org/wiki/Free_and_open-source_software.

———, Free software, http://en.wikipedia.org/wiki/Free_software.

———, General circulation model, http://en.wikipedia.org/wiki/Global_climate_model.

———, Gratis versus libre, http://en.wikipedia.org/wiki/Gratis_versus_libre.

———, GRIB, http://en.wikipedia.org/wiki/GRIB.

———, IT service management, http://en.wikipedia.org/wiki/IT_service_management.

———, Meteorological reanalysis, http://en.wikipedia.org/wiki/Meteorological_reanalysis.

———, NetCDF, http://en.wikipedia.org/wiki/NetCDF.

———, Non-functional requirement, https://en.wikipedia.org/wiki/Non-functional_requirement.

———, OPeNDAP, http://en.wikipedia.org/wiki/OPeNDAP.

———, Open data, http://en.wikipedia.org/wiki/Open_data.

———, Solution architecture, http://en.wikipedia.org/wiki/Solution_architecture.

———, Spatial data infrastructure, http://en.wikipedia.org/wiki/Spatial_data_infrastructure.

———, Swim lane, http://en.wikipedia.org/wiki/Swim_lane.

———, Teleconnection, http://en.wikipedia.org/wiki/Teleconnection.

———, Uncertain data, http://en.wikipedia.org/wiki/Uncertain_data.

———, Uncertainty, http://en.wikipedia.org/wiki/Uncertainty.

———, Use case, http://en.wikipedia.org/wiki/Use_case.

———, Web Processing Service, http://en.wikipedia.org/wiki/Web_Processing_Service.

Appendix 1

Diagrams of climate data management system components
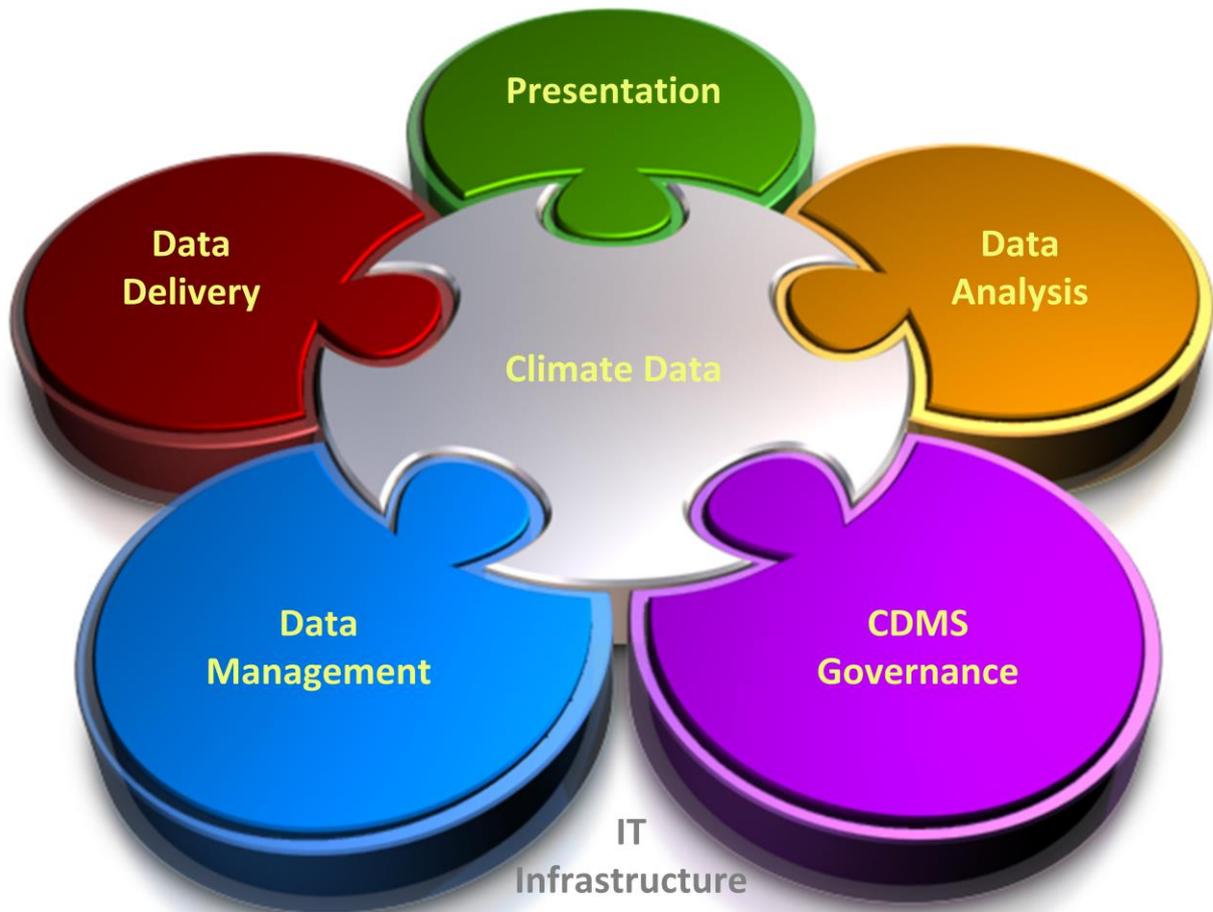
High Level Components of a CDMS

Overview of CDMS Functional Components

Climate Data Presentation

Printed Media

Written Reports

Graphical User Interface – Time Series Data Exploration

Business Intelligence Reports | Manage Content | Visualisation | Integrated Search of Climate Data | Data Download

Climate Data Delivery Services

Open Spatial Standards

OGC Services | GML Application Schema

Data Discovery

Climate Metadata Registers

Other Formats | WMO Formats

Climate Data Analysis

Analysis

Climate Modelling | Generate Derived Data from Climate Observations | Data Homogenisation | Other

Climate Data Management

Extract Transform Load

Data Extraction | Data Ingest

Data Rescue

Imaging | Data Entry | Monitoring

Observations Quality Control

Quality Management | Data Monitoring | Network Monitoring

Quality Assessment

Observations Quality Assessment | Uncertainty | Derived Data QA

Climate Metadata

Manage Climate Metadata

Time Series Climate Data

Climate Metadata

Observations Metadata | Data Provenance | Dataset Discovery Metadata

Observations Data | **Logical Data Models** | Climate Database | Foundation Standards | WMO Logical Data Models

WMO Standard Products

Observation Data Products | Climate Change Indices

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 Management

Computing Infrastructure

Networks | Computing Platform | Security | Data Storage | Disaster Recover

CDMS Governance

Data Policy

Commitments | Sustainability | Intellectual Property | Data Delivery | Third Party Data | Climatology Policy

Governance

Data Governance | IT Governance

Training | System enhancement capability

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

137

CDMS Specifications

CDMS Governance

Data Policy

Commitments
- WMO Resolutions
- WMO Technical Regulations
- WMO Technical Commissions
- International
- National

Sustainability
- Funding
- Data Custodian
- Access to Data
- Archival
- Disaster Recovery

Climatology Policy
- Data Lineage Traceability
- Climate Networks
- Data Generation
- Quality Assurance
- Future Climate Data Framework
- Climate Metadata
- Sensor or Station Change

Governance

Data Governance
- Controlled Access to Data and Systems
- Approval Process New Data Type
- Approval Process to change data
- IT Change Approvals - No Data Corruption

IT Governance
- IT Service Management
- Managed Change
- Project Management
- IT Architecture
- Documentation

Intellectual Property
- Data Licencing
- Access Constraints
- Copyright
- Usage Constraints
- Attribution

Data Delivery
- Interoperability Standards
- Cost Recovery
- Data Quality Delivered

Third Party Data
- Crowd Sourced
- Other Agency
- Commercial

Required | Recommended | Optional

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

CDMS Specifications

Time Series Climate Data

Climate Metadata

Observations Metadata
- Station Identifier
- Station Overview
- Location
- Sensor
- Station Status
- Station Type
- Network
- Data Processing
- Local Environ
- Data Transmission

Data Provenance
- What was changed
- What was done to change it
- Who did they act on behalf of
- When was it changed
- How / Why was it changed
- What was it derived from
- Who / What Changed it
- Who was responsible

Dataset Discovery Metadata
- Dataset Identifier
- Dataset Overview
- Dataset Data Quality
- Distribution
- Access Constraints
- Dataset Maintenance
- Spatial Representation
- Reference Systems

Observations Data

Climate Observations
- Atmospheric
- Terrestrial
- Oceanic

Logical Data Models

Climate Database
- Data Dictionary

Foundation Standards
- Observations and Measurements

WMO Logical Data Models
- METCE
- Climate Observations Application Schema

WMO Standard Products

Observation Data Products
- Routine Messages
- Climatological Standard Normals
- World Weather Record
- CLIMAT
- Aeronautical Climatology
- Other

Climate Change Indices
- Core Indices
- Other Indices

Derived Climate Data

Derived Observation Data
- Homogenised Data
- Computed
- Normals and Averages
- Other

Gridded Spatial Distribution
- Analysed Data
- Other

Numerical Models
- Numerical Models

Ancillary Data

Spatial
- Topography
- Impacts Data
- Administrative
- Agricultural
- Other
- Health
- Environmental
- Emergency Management

Climate Documentation
- Documentation
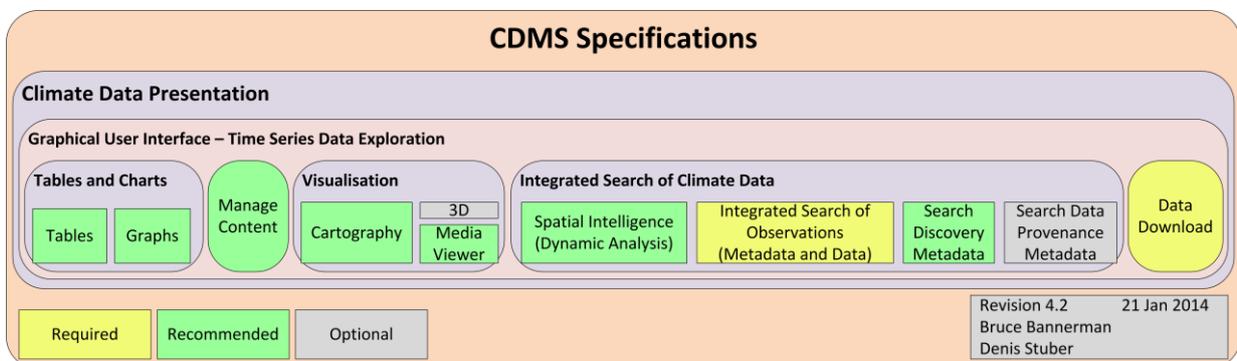- Various Media
- Published Reports

Climate Software
- Source Code
- Testing
- Package Management
- Environment Configuration

Required | Recommended | Optional

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

CDMS Specifications

Climate Data Management

Ingest and Extract

Data Ingest

- WMO Messages
- Business Rules
- Vector
- Status Log
- Raster array
- Automated with self-recovery
- Other formats
- Transform
- Data Extraction

Data Rescue

Imaging

- Document Imaging
- Optical Character Recognition
- Chart Digitisation

Data Entry

- Forms
- Key Entry
- Computation

Monitoring

- Data Rescue metrics

Observations Quality Control

Quality Management

- Consistancy Checks
- Statistical Checks
- Heuristic Checks
- Spatial Checks
- Data Comparison
- Data Recovery
- Data Monitoring
- Network Monitoring

Quality Assessment

Uncertainty

- Derived Data
- Measurements

Observations Quality Assessment

- Siting Classification
- Multi-Layer Quality Flags
- Sustained Performance Classification
- Climate Observation Quality Classification
- Derived Data Quality Assessment
- Quality Assurance Metrics

Climate Metadata

Manage Climate Metadata

- Create
- Maintain
- Quality Control
- Metrics

Legend: Required | Recommended | Optional

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

CDMS Specifications

Climate Data Analysis

Analysis

Climate Modelling

- Numerical Models
- Reanalysis
- Model Ensembles

Generate Derived Data from Climate Observations

- Spatial Analysis
- Image Analysis
- Time Series Analysis
- Teleconnection Indices

- Data Homogenisation
- Other

Legend: Required | Recommended | Optional

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

CDMS Specifications

Climate Data Presentation

Graphical User Interface – Time Series Data Exploration

Tables and Charts

- Tables
- Graphs

- Manage Content

Visualisation

- Cartography
- 3D
- Media Viewer

Integrated Search of Climate Data

- Spatial Intelligence (Dynamic Analysis)
- Integrated Search of Observations (Metadata and Data)
- Search Discovery Metadata
- Search Data Provenance Metadata

- Data Download

Legend: Required | Recommended | Optional

Revision 4.2   21 Jan 2014
Bruce Bannerman
Denis Stuber

CDMS Specifications

Climate Data Delivery Services

Open Spatial Standards

OGC Services

Web Map Services	Web Feature Services	Web Coverage Services	Sensor Web Enablement Services	Web Processing Services
CF netCDF	GeoSynchronisation		Symbology Encoding	Styled Layer Descriptors

GML Application Schema

Combined Climate Observations and Metadata

Taxonomies and Registers of Authoritative Terms

Data Discovery

Catalogues

Discovery Metadata

Observations Metadata

Provenance Metadata

Linked Data

Other Formats

OpenDAP

WMO Formats

WMO Formats

Required	Recommended	Optional

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

CDMS Specifications

Core IT Infrastructure

Application Infrastructure

Identity Management

Directory	Identity and Access Management

Collaboration

Email	ftp	wiki

Web Platform

Web Server	Proxy Server

Database

Tabular	Spatial

Service Operations

Service Operations Management

Scheduling	Service Desk

Applications Management

System Management

Computing Infrastructure

Networks

Internet	WMO WIS/GTS	Internal Networks	VPN

Computing Platform

Hardware	Operating System	High Performance Computing

Security

Data Storage

Storage Media	Data Archival	Backups

Disaster Recover

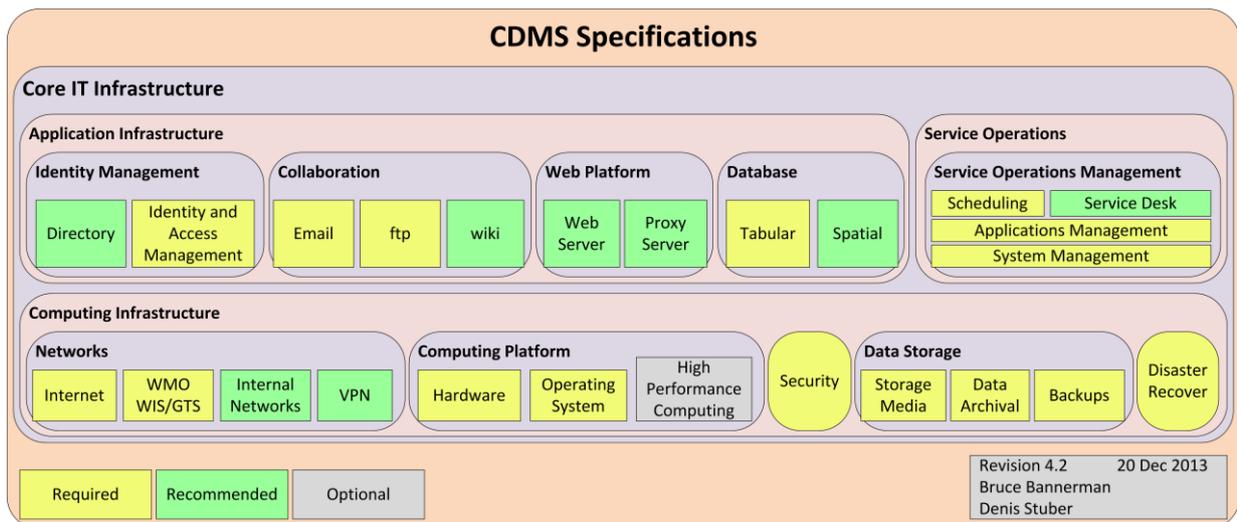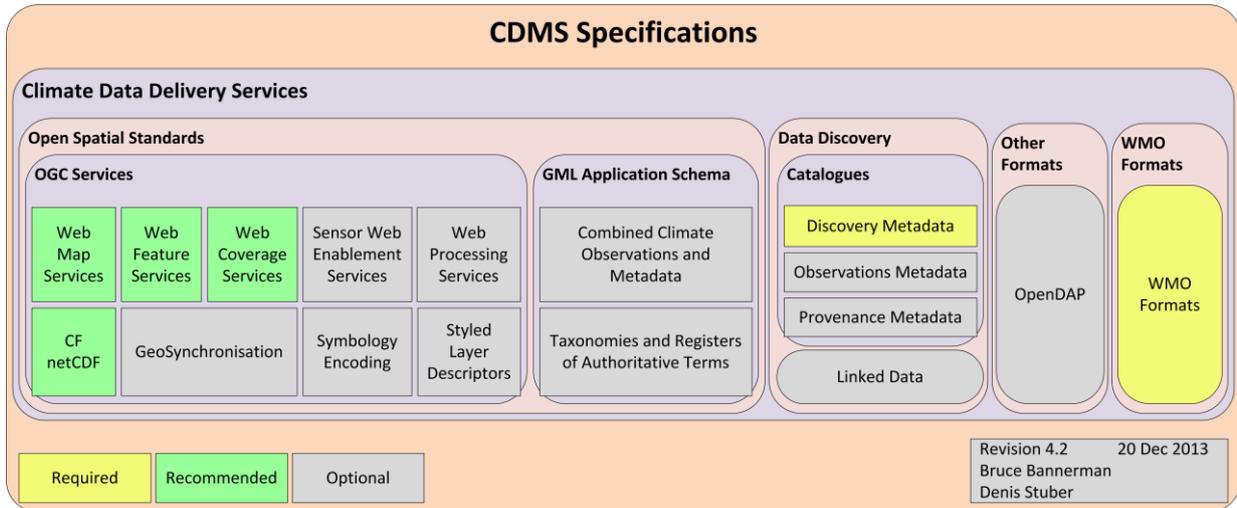| Required	Recommended	Optional

Revision 4.2   20 Dec 2013
Bruce Bannerman
Denis Stuber

140

Appendix 2

List of required components for climate data management systems

List of required components

Chapter	Section	Subsection	Component	Component number
CDMS governance				**3**
	Data policy			3.1
		Commitments		3.1.1
			WMO resolutions	3.1.1.1
			WMO Technical Regulations	3.1.1.2
			WMO technical commission guides	3.1.1.3
			International	3.1.1.4
			National	3.1.1.5
		Sustainability		3.1.2
			Disaster recovery	3.1.2.1
			Funding	3.1.2.2
			Data custodian	3.1.2.3
			Access to data	3.1.2.4
			Archival policy	3.1.2.5
		Intellectual property		3.1.3
			Data licensing	3.1.3.1
		Data delivery		3.1.4
			Quality of delivered data	3.1.4.2
		Climatology policy		3.1.6
			Climate metadata	3.1.6.1

Data rescue		5.2
	Data entry	5.2.3
		Forms	5.2.3.1
		Key entry	5.2.3.2
Observations quality control		5.3
	Quality management	5.3.1
		Consistency checks	5.3.1.1
		Heuristic checks	5.3.1.3
		Statistical checks	5.3.1.4
		Data recovery	5.3.1.6
Quality assessment		5.4
	Observations quality assessment	5.4.1
		Siting classification	5.4.1.1
		Multilayer quality flags	5.4.1.3
	Uncertainty	5.4.4
		Measurements	5.4.4.1
Climate data presentation		7
	Graphical user interface – time-series data exploration	7.1
	Integrated search of climate data	7.1.4
		Integrated search of observations (metadata and data)	7.1.4.2
	Data download	7.1.5
		Data download	7.1.5.1
Climate data delivery services		8
	Data discovery	8.2
	Catalogues	8.2.1
		Discovery metadata catalogue	8.2.1.1
	Other formats	8.3
	WMO formats	8.3.2

	WMO formats	8.3.2.1
Core IT infrastructure		**9**
Application infrastructure		9.1
	Identity management	9.1.1
	Identity and access management	9.1.1.2
	Collaboration	9.1.2
	E-mail	9.1.2.1
	FTP	9.1.2.2
	Database	9.1.4
	Tabular	9.1.4.1
Service operations		9.2
	Service operations management	9.2.1
	Scheduling	9.2.1.1
	Applications management	9.2.1.3
	Systems management	9.2.1.4
Computing infrastructure		9.3
	Networks	9.3.1
	Internet	9.3.1.1
	WMO WIS/GTS	9.3.1.2
	Computing platform	9.3.2
	Hardware	9.3.2.1
	Operating system	9.3.2.2
	Security	9.3.3
	Security	9.3.3.1
	Data storage	9.3.4
	Storage media	9.3.4.1
	Data archival	9.3.4.2
	Backups	9.3.4.3

Disaster recovery	9.3.5
Disaster recovery	9.3.5.1