WMO Strategy to Improve Global CDMS Implementation 2018 - 2023

An important point to remember is that the most important stakeholders for our data have not been born yet and that we are custodians of this data for their future use.[1]

[1] Statement attributed to Dr Blair Trewin, Australian Bureau of Meteorology

Document Name:	WMO Strategy to Improve CDMS Implementation
Author:	Bruce Bannerman, Co-Chair ET-CDMS

Document Sponsor:	WMO Commission for Climatology
Document Version:	0.11
Document Status:	Draft
Approval Status:	Not Yet Approved

Table of Contents

Executive Summary

A CDMS is an essential tool required to support the management, analysis and delivery of climate data within NMHS, regionally and globally.

Climate data allows a scientist to understand for a given location, both the meteorological conditions related to a specific event; and the climatic conditions that have occurred, or may occur, over an extended period of time.

Climate data is required routinely for a wide range of purposes. It is an essential input that is used in a very wide range of decisions that are made routinely to support society's needs. This includes decisions related to the GFCS Priority Areas of: Agriculture and food security; Disaster risk reduction; Energy; Health; and Water.

Modern CDMS as defined by WMO #1131, Climate Data Management System Specifications are intended to provide a fundamental capability that **to underpin** Global Framework for Climate Services (GFCS) requirements. These include at least:

- Storage, management, quality control and exchange of climate observations required by the Observations and Monitoring component of the GFCS; and
- A number of functions required by the Climate Services Information System (CSIS) component including:
    - Standardised management and exchange of climate and climate-related data as per WMO resolutions;
    - Monitoring and analysing climate variability on different temporal scales, including extremes such as droughts and floods;
    - Deriving products (datasets, text, maps, charts, statistics, etc.) that describe the past, present and future climate of a location, country, region and indeed the whole globe; and
    - Deriving tailored products and information within a range of social, economic and environmental contexts based on the tools and guidance developed by the User Interface Platform.

There are significant issues globally with the management and exchange of climate data. Many of these issues relate to the lack of suitable CDMS capabilities within NMHS. The CDMS that do exist, typically do not comply with modern requirements as defined by WMO #1131.

This strategy aims to make a substantial improvement in the availability of climate data globally. It is intended to:

- Establish solid climate data management foundations that will be required by future climate data services; and
- Significantly improve the availability of modern CDMS within a sustainable, coordinated framework. Priority will be given to the needs of developing countries.

Vision of this CDMS Strategy - 2018-2023

Achieve a step-change in the management of climate data by:

- Thinking globally of what climate data management capability is required to answer many questions of societal need; and
- Acting nationally to implement and evolve consistent and sustainable CDMS that will address national requirements for climate data, and in addition help to address regional and global needs.

10+ year CDMS Vision

Federated global data services are a reality for many requirements, e.g. consider this user story:

> *A climate researcher wishes to undertake an annual climate forecast analysis. The climate researcher initiates an automated process that downloads observations data and associated observations metadata for all available countries in the world. All data conforms to the WMO climate observations logical data model, with a known structure and semantics. The data is automatically loaded into the researcher's CDMS. The researcher undertakes the required analysis without having to waste considerable time worrying about the semantics and formatting of the source data.*

Background

What is a Climate Data?

Definition

Climate data allows a scientist to understand for a given location[2], both:

- The meteorological conditions related to a specific event; and
- The climatic conditions that have occurred, or may occur, over an extended period of time.

For a more in-depth architectural overview of climate data, please see Appendix A.

What is a CDMS?

Definition

A CDMS is an essential tool required to generate, manage and deliver high quality climate data that is in turn required as input to a wide range of decisions that are made for societal benefit.

More formally: *a Climate Data Management System (CDMS) is an integrated computer-based system that facilitates the effective archival, management, analysis, delivery and utilisation of a wide range of integrated climate data*.

Current use of the term is inconsistent

Current software systems that are used to manage climate data are often labeled as a 'CDMS'. However, all systems that are currently in use have been developed in the absence of consistent specifications.

Therefore each current 'CDMS' has different and inconsistent capabilities. It is very difficult to currently compare the functionality of, or the data from, two or more CDMS.

Approved Climate Data Management System Specifications

Clear and consistent CDMS guidance was introduced in July 2014, with the publishing of **WMO #1131, Climate Data Management Systems Specification**. WMO Congress approved WMO #1131 as a WMO Technical Specification in 2015 (Cg-17, p463).

WMO #1131 has been designed to address a range of modern information management challenges, take into account WMO strategic directions, and to provide a clear definition of the functionality expected within a CDMS.

The CDMS Specifications are designed as modular components to allow the greatest flexibility for CDMS development, maintenance, deployment and upgrade.

Readers are advised to drop any assumptions that they have regarding what a CDMS is and to review WMO #1131 for an understanding of the functionality that is now expected in modern CDMS.

[2] The term location is used here in a very broad sense. A location could be a site, local area, country, region, continent, or the whole world.

High Level Components of a CDMS

Figure 1 High Level Components required in a modern CDMS

Drivers for change

Who needs to use high quality climate data?

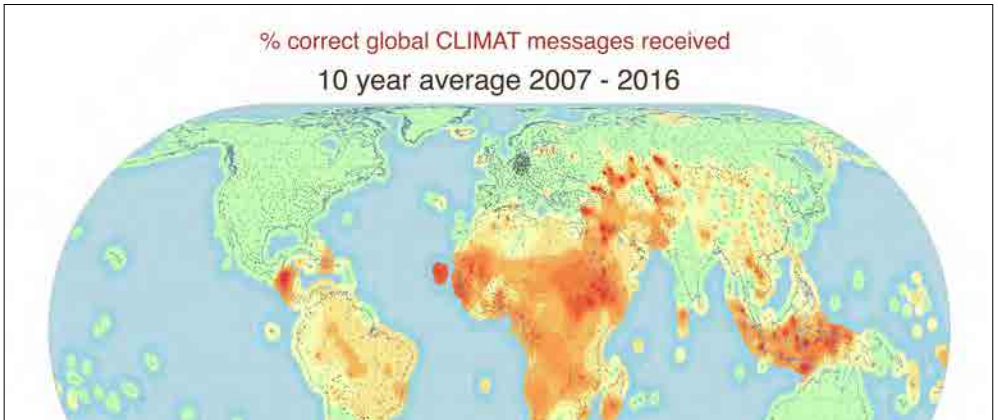
High quality climate data is required all the time. It is an essential input that is used in a very wide range of decisions that are made routinely to support society's needs.

Some examples are:

- A water manager can decide to enforce water restrictions based on the climate forecast for the coming season and information on historical rainfall over a catchment.
- A farmer can weigh the risks of planting certain seeds and crops for the coming season using the El Niño climate forecast and historical information about his property.
- An epidemiologist can provide early warning on the increased risk of malaria by analyzing the relationship between disease outbreaks and prevailing climatic conditions over a region.
- An emergency services analyst can plan bushfire risk mitigation actions for the coming season by investigating the relationship between vegetation curing, and forecast and historical climatic conditions.
- A renewable energy analyst can select the best type of solar system for a client by reviewing historical solar radiation data for a region.
- A civil engineer can determine the optimal location for a road and bridge project by analyzing the historical potential for rain and floods in particular areas.
- A land manager can create a plan to mitigate climate change impacts on biodiversity and ecosystems by analyzing climate change projections and historical climatic conditions for certain species.
- A city planner can consider whether a housing development proposal should be approved using information on the historical potential for rain and floods in the area.

What is the current global experience of reliable climate data delivery?

Using an example of the delivery of one of the basic climate data types, the monthly CLIMAT message, an analysis of the delivery of valid CLIMAT messages over the decade from 2007 to 2016 was undertaken in mid-2017.



This analysis indicates that there are fundamental global issues with the management and delivery of climate data.

What is the current global experience of CDMS implementation?

When the results of the ten year valid CLIMAT message analysis is combined with the CDMS Survey that was conducted in 2012 by the WMO Expert Team – Climate Data Management Systems (ET-CDMS), the global situation becomes more readily apparent. While this analysis is five years old, it is the opinion of the ET-CDMS Co-Chairs that it is still indicative of the current global CDMS situation. Follow up surveys are planned.

The survey was sent to 191 WMO Members, of which 72% responded. A summary of the findings of the 2012 CDMS survey include:

Do not use a database

Use a spreadsheet to manage climate data

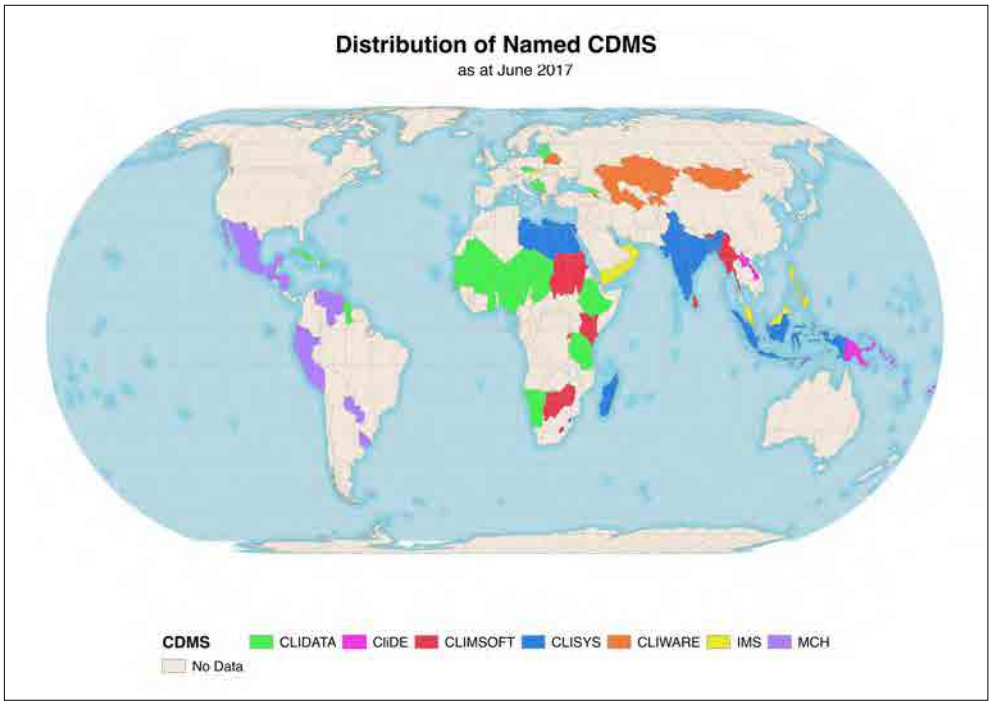
Have developed their own CDMS

Have operational issues

CDMS is not operational

Would like to change their CDMS

To further illustrate this issue, ET-CDMS undertook work in early 2017 to understand the global distribution of ‘named’ CDMS, sometimes referred to as Franchise CDMS. The remaining countries either use no CDMS, or use CDMS that they have developed themselves.



Each of these ‘named’ CDMS is developed under short term intermittent project funding, by only one or two people. There are limited funds available for CDMS maintenance. This situation has resulted in key person dependencies.

At least two of the main open source CDMS used have lost key staff due to short term project funding, combined with future funding uncertainties. This has resulted in significant impact on project viability and progress when funds again became available.

In addition, on investigation, it was found that there are **42 developing countries** that rely on such open source CDMS that are at risk of failure through inadequate funding.

It is estimated that for approximately 190 WMO members, there are 96 different CDMS. Given the changes required to meet modern CDMS requirements, the costs for developing and maintaining CDMS, together with the shortage of funds available **this situation is clearly unsustainable**.

Themes of this Strategy

This strategy outlines a response to address and significantly improve the current CDMS situation within a sustainable framework.

Noting that the vision of the CDMS Implementation Strategy 2018 - 2023 is:

> *Achieve a step-change in the management of climate data by:*
>
> - *Thinking globally of what climate data management capability is required to answer many questions of societal need; and*
> - *Acting nationally to implement and evolve consistent and sustainable CDMS that will address national requirements for climate data, and in addition help to address regional and global needs.*

The key strategic actions to achieve this vision of the WMO Strategy to improve global CDMS implementation 2018-2023 are outlined below. These actions include:

- Find a 'home' for CDMS activities;
- Establish a robust global framework for CDMS related activities to operate within;
- Further develop formal CDMS specifications;
- Establish consistent community agreed formal data definitions;
- Establish clear, unambiguous metrics to measure improvements; and
- CDMS Rationalisation;

Thinking globally

From the CDMS Strategy Vision:

> *Achieve a step-change in the management of climate data by:*
> *Thinking globally of what climate data management capability is required to answer many questions of societal need.*

The *'thinking globally'* aspects of this strategy are intended to establish solid foundations for climate data management, globally. These foundations will be required to underpin future climate data services, particularly as WIS-2 evolves. As an example, these solid foundations, together with WIS-2 will be required to support the 10+ year CDMS vision described at page 6.

There are a number of activities that are addressed as part of *'thinking globally'*. These include:

- Find a 'home' for CDMS activities;
- Establishing a robust global framework for CDMS to operate within;
- Establish clear unambiguous metrics to measure improvements; and
- CDMS rationalisation.

These topics are discussed in more detail below.

Find a 'home' for CDMS activities

To realise the potential outlined in this strategy, a 'home' is required for CDMS related issues.

It is recommended that this 'home' is assigned the mandate, responsibility and accountability to oversee and coordinate CDMS related requirements and activities.

This will in turn provide a much-needed focus to obtain necessary funding and to undertake sustained global CDMS related activities.

Three factors for consideration in selecting a 'home' for CDMS activities are:

Commission of Climatology

The Commission of Climatology (CCl) has been very supportive of CDMS activities over the last decade. Through the CCl stewardship, WMO now has:

- An approved WMO Technical Specification that defines the functionality expected in modern CDMS (WMO #1131);
- Plans for a modern open source Reference CDMS;
- A clear understanding of the current state of global CDMS deployments;
- An approved Data Rescue and CDMS Resource plan; and
- This CDMS Implementation Strategy – 2018-2023.

WMO Restructure Plans

Given the current restructure plans within WMO, it is unclear as to whether CCl will continue to exist in the future. If it doesn't, an alternate appropriate 'home' for CDMS activities may become readily apparent; and

The Global Framework for Climate Services

Modern CDMS as defined by WMO #1131, Climate Data Management System Specifications aim to provide a fundamental capability that is required to underpin Global Framework for Climate Services (GFCS) requirements. These include at least:

- Storage, management, quality control and exchange of climate observations required by the Observations and Monitoring component of the GFCS; and
- A number of functions required by the Climate Services Information System (CSIS) component including:
    - Standardised management and exchange of climate and climate-related data as per WMO resolutions;
    - Monitoring and analysing climate variability on different temporal scales, including extremes such as droughts and floods;
    - Deriving products (datasets, text, maps, charts, statistics, etc.) that describe the past, present and future climate of a location, country, region and indeed the whole globe; and

- Deriving tailored products and information within a range of social, economic and environmental contexts based on the tools and guidance developed by the User Interface Platform.

Data managed within CDMS are required as input into a range of processes and analysis that is undertaken within the GFCS Priority Areas of: Agriculture and food security; Disaster risk reduction; Energy; Health; and Water.

This CDMS Implementation Strategy conforms to a number of the GFCS Principles including:

- ***High priority for the needs of climate-vulnerable developing countries;***
  - This strategy prioritises the deployment of essential climate data management capability to developing countries within a coordinated and sustainable framework;
  - The strategy aims to implement basic capability that is readily available now for immediate benefit and facilitate scheduled upgrades to the implemented CDMS when appropriate; and
  - Concurrently, guide the upgrade of CDMS to meet modern requirements as defined by WMO #1131, while rationalising the number of available CDMS into a number that is more affordable to maintain and develop sustainably for the long term.
- ***Primary focus is the better access and use of climate information by users;***
  - By implementing sustainable CDMS to many countries in a coordinated framework, this strategy will ensure that the necessary technical tools are available to make it easier to capture, manage, use, analyse and access climate information for a wide range of uses;
  - In addition, this strategy prepares solid foundations that will be required by modern climate data services that operate within a future WIS-2 environment.
- ***Framework will address needs at three spatial scales: global, regional and national;***
  - At a high level, this strategy is focused on:
    - implementing and improving climate data management capability at a national level, that is coordinated regionally within Regional Climate Centres and globally by an entity that has yet to be decided;
    - while thinking about what is required to facilitate global improvements in the management of climate data; and
    - these global requirements will in turn be built into future CDMS to provide the same capabilities to each country. These capabilities will then be implemented nationally via coordinated upgrades, to facilitate a step-change in the global management of climate data.
- ***Climate services must be operational and continuously updated;***
  - This strategy aims to implement operational CDMS within NMHS in a coordinated and sustainable manner. This includes the maintenance of the CDMS software to ensure long term stability; and
  - The CDMS is capable of ongoing updates of data to underpin a range of climate services that require dynamic data sources.

When the above is considered, the GFCS may also be a good 'home' for CDMS activities.

Establish a robust global framework for CDMS related activities to operate within

Noting:

- Recommendation 11.1 of WMO #1131, CDMS Specifications pp 142-146, Establish the Climate Data Framework.
- Recommendation 11.2 of WMO #1131, CDMS Specifications pp 146-147, Continue developing CDMS Specifications.
- Recommendation 11.5 of WMO #1131, CDMS Specifications p 151, Fast track development of WMO logical data model.
- That WMO #1131 was approved by Cg-17.

There is a range of activities that will need to be addressed to provide a robust global framework for CDMS related activities to operate within.

Where possible, it is preferable to utilise existing arrangements, rather that creating yet another arrangement for CDMS.

Governance

The governance surrounding CDMS activities will need to be clearly defined. There are several aspects to establishing a suitable governance regime for CDMS activities. These include:

- There needs to be a clear understanding of what decision making process is required to support CDMS related issues, and who is accountable for the decisions. See the above section on "Find a 'home' for CDMS activities";
- There will be different specialisations, levels of responsibility and accountability to be taken into account. Some are discussed below.
- In addition, there are many coordination activities as discussed at Recommendation 11.1 of WMO #1131, CDMS Specifications pp 142-146, Establish the Climate Data Framework. This activity, currently named a Climate Data Framework, relates to defining the technical environment, together with the policies, rules and institutional arrangements that will be required to support future dynamic and federated climate data services as touched on in the 10+ year CDMS Vision on page 6 of this strategy. A description of these requirements is provided at Recommendation 11.1 of WMO #1131
- Climate Data Framework activities go well beyond what is currently being addressed in IPET-CDMP and will overlap with what will need to be developed to facilitate WIS-2.

Recommendations:

- Clearly define CDMS related Governance and decision making processes; and
- Coordinate work within both the WIS-2 and IPET-CDMP development processes to ensure that Climate Data Framework outcomes are met without the need to establish yet another framework that is dedicated to climate data requirements.

Project Governance

The project activities outlined in this strategy will require substantial funding and project governance to ensure that the potential described can be achieved.

Recommendation:

- It may be appropriate to extend an existing framework such as GFCS to cover CDMS project requirements and activities.

Technical Governance

Technical governance is expected to be problematic;

In the long term, CDMS will need to comply with, create new and possibly revise a range of WMO related technical requirements. Some examples are GFCS, CSIS, WIS-2, WMO Codes Registry, WIGOS, HQDMF for Climate, etc.;

A number of these requirements are either just starting development, or still a work in progress;

It is expected that CDMS developers will not be able to stop work on specific issues indefinitely while issues are resolved via often lengthy WMO processes;

Therefore it will be prudent to allow CDMS development to progress as required with minimal overhead; and

A process will need to be developed that will allow interim solutions to be arrived at that can be used by CDMS developers. This will be on the understanding that when appropriate, CDMS will need to be transitioned to comply with official WMO guidance when it is ready, and work scheduled accordingly within CDMS project(s).

Recommendations:

- Coordination of issues that affect CDMS within and between these formal WMO programs will need to be undertaken, with appropriate mandate and accountability assigned as appropriate. CDMS related contribution to and feedback on the development of these initiatives is required.
- It is proposed that CDMS Developer meetings arrive at and maintain an appropriate interim process that is to be used by all CDMS development projects.

Further develop formal CDMS specifications

As requested by Congress (Cg 17, p463) further work is required to extend WMO #1131 with a view to creating WMO standards for climate data management and related systems.

Some examples of the work required are:

- Changes for clarification and to reduce WMO duplication (e.g. liaison with TT-IM, TT-WMD);
- Extend work on user requirements, verification that other WMO Expert Teams needs are met;
- Continue development of conceptual architecture to support developers; and
- Establish a process for approving peer-reviewed algorithms.

Recommendation:

- Therefore it will be necessary to continue the work undertaken by CCl's Expert Team – Climate Data Management Systems.

Establish consistent community agreed formal data definitions

Formal data definitions, data models and data products that are community agreed are required to support future advanced climate data services.

Too much time is wasted on data integration issues when undertaking regional and global climate data analysis. Much of this time is spent on trying to get poorly defined data products into a form that will support consistent analysis. There have been estimates that this activity may take up to 80% or more of an analysis project.

This issue needs to be addressed, particularly if the vision of federated global climate data services is to be achieved in ten years.

Work is required to establish consistent community agreed formal data definitions, data models and data products.

These formal data definitions can then be built into the capabilities of CDMS and implemented widely in order to:

- Facilitate ease of data exchange;
- Minimise effort required to aggregate data for global analysis; and
- Facilitate dynamic views of data as globally federated CSIS data services.

This activity will take many years to undertake.

WMO has a good head start with the work undertaken on the GCOS Essential Climate Variables, however the current work is insufficient to support formal data definitions.

The work undertaken within the Open Geospatial Consortium on the development of the Observations and Measurements, WaterML2 and Timeseries standards is more like what is required.

WMO has already undertaken some work on progressing such standards. Some examples are:

- work on a common base data model in the development of the METCE data standard;
- supported by the WMO Codes Register; and
- Initiation and supporting work on the OGC Timeseries standards.

Much of this work was coordinated within the Commission for Basic System (CBS) Inter-Program Expert Team – Metadata and Data Representation Development (IPET-MDRD) in the last Intercessional Period.

Recommendations:

- Formal definitions are required for climate data. It is recommended that the work on METCE and the WMO Codes Register be re-started with a view to extending the concepts to cover at least the GCOS Essential Climate Variables;
- Given the amount of work required and the pressing need, find a more sustainable way of undertaking this activity. Relying on voluntary contributions within a WMO Expert Team scenario will not achieve the results in the timeframe required; and
- Consider utilising the Memorandum of Understanding that exists between WMO and the Open Geospatial Consortium (OGC) and working within OGC Technical Committee processes to establish the required climate data standards. OGC processes allow for

wide stakeholder engagement and input. They have proven effective in the development of data standards like WaterML 2. However, the work will need to be sponsored and coordinated by WMO with dedicated resources allocated to achieve the required aims.

Establish clear, unambiguous metrics to measure improvements

As there may be considerable funds spent on improving the global CDMS situation, it is very important to establish and maintain suitable metrics that can be used to assess the effectiveness of the investment.

These metrics will provide project boards with a means of assessing performance and if required discontinuing poor performing activities.

understand CDMS baseline

In order to understand the baseline that future CDMS activities can use to measure performance and improvements, the following metrics can be used:

- 10 year CLIMAT analysis – 2007-2016
- 2012 CDMS survey

Recommendations:

- It is recommended that the following additional maturity model and metrics be established **as a priority** to help establish a set of CDMS baseline metrics:
    - A CDMS maturity model and self-assessment tool that can be used by both NMHS and CDMS Developers to understand their current CDMS situation. These tools can also be used to support arguments for future CDMS investment within NMHS; and
    - An index showing country compliance with Resolution 60 (Cg17).

measure performance and monitor improvements

In order to understand relative improvements against the CDMS baseline, the following metrics should also be established and used to provide objective metrics to determine if CDMS investment is achieving the desired goals:

- Scheduled Global CLIMAT analysis;
- Scheduled CDMS Surveys;
- Scheduled CDMS maturity model Surveys;
- Scheduled surveys to show improvements in countries that provide data in compliance with Resolution 60 (Cg17);
- Scheduled surveys to show improvements in the the availability of NCMP (National Climate Monitoring Products);
- A Register of CDMS, together with a summary of their capabilities and their compliance with WMO #1131; and
- Other indices as deemed appropriate.

Recommendation:

- Establish and maintain additional CDMS related metrics as required.

CDMS Rationalisation

Noting:

- The findings of the 2012 CDMS Survey that there are:
    - Significant issues with the implementation of CDMS globally; and
    - Estimated to be 96 CDMS for 190 countries and that these CDMS provide different and inconsistent capabilities.
- ET-CDMS work from early 2017 that shows that 42 developing countries are reliant on open source CDMS that are under severe financial constraints, have single points of failure in key developers, and are consequently at risk of failure;
- Recommendation 11.4 of WMO #1131, CDMS Specifications pp 148-150. This recommendation was for WMO to sponsor and support open source approaches to CDMS development.
- WMO #1131 was approved by Cg-17.
- ET-CDMS Policy Statement on CDMS dated November 2015[3], approved by CCI Management Group.
- The ten year CLIMAT analysis 2007-2016 that indicates significant issues with climate data management globally.

Background to CDMS Rationalisation

Please see Appendix B, CDMS Rationalisation for a more in-depth discussion on:

- Why this activity is required;
- The expected benefits to the WMO Community, NMHS and Climate Science; and
- Suggested High Level Rationalisation goals for the short, medium and long term.

Appendix B also includes the CDMS Rationalisation text that is included in this section of the Strategy, but with more context.

There are estimated to be 96 different CDMS in use globally. These CDMS have been developed in the absence of CDMS Specifications and it is understood that none have the functionality expected in a modern CDMS as defined in WMO #1131.

It is understood that all CMDS developers are facing severe constraints in obtaining sufficient funds to develop and maintain their systems.

It will be a waste of scarce funds and resources to attempt to develop the same capability within each of the 96 CDMS in order to upgrade them with the functionality that is expected within a modern CDMS.

It is time for a rethink on how CDMS are developed.

This CDMS Strategy proposes that NMHS and other interested parties **collaborate and pool funds** to develop and maintain a single Reference CDMS that is designed to provide the functionality expected within a modern CDMS as defined by WMO #1131.

This Reference CDMS is nominally called Open-CDMS as at the date of this strategy document.

[3] http://wis.wmo.int/file=2253

This Reference CDMS would be developed and maintained using tried and tested open source software development practices within an open source software community.

For an overview of why an open source approach has been selected for this CDMS Implementation Strategy, see WMO #1131, Recommendation 11.4, pp148-150.

This approach will reduce the long-term financial outlay that WMO, NMHS, and other organisations expend to develop and maintain various CDMS.

WMO #1131 is designed as a set of inter-related components that can be combined to provide the capability expected within a modern CDMS. Many of these components describe functionality that is often readily available within existing open source and proprietary software packages.

Therefore it is not expected that the Reference CDMS will be a single monolithic software application, but a collection of software applications that are developed, extended, integrated and bundled to provide the required CDMS functionality.

This approach will allow for future innovation within software components, allowing individual components to be upgraded and replaced, rather than replacing the complete CDMS.

Similarly this approach will allow the Reference CDMS to be scaled to meet a range of needs, from the needs:

- Of developing countries; to
- The advanced technology needs of many NMHS, Regional Climate Centres and Global Climate Centres.

The work in designing, developing, integrating and upgrading the Reference CDMS will need to be funded and coordinated.

This CDMS Strategy expects that this Reference CDMS work will be undertaken within an open source software community and that any software development is undertaken as an open source software project. This approach will:

- Establish and provide the rules and governance on how CDMS work is undertaken;
- Allow collaboration by any interested party, including: UN; WMO; NMHS; proprietary and open-source commercial organisations; open source software developers and communities; and members of acadaemia.
- Spread the cost and burden of developing the Reference CDMS across many organisations;
- Ensure that the results of any effort is not locked up in closed source, black-box proprietary solutions and that the Intellectual Property and results of any investment can be re-used in the future with little or no constraints;
- Ensure that the software developed to manage, analyse and provide climate data is open for review and scrutiny;
- Facilitate a transfer of knowledge for those interested in learning more about CDMS development; and
- Develop a robust and collaborative Open Source Software community to facilitate the long-term viability and sustainability of the Reference CDMS.

The Reference CDMS is nominally called Open-CDMS. This name may be reviewed in the future.

Priority in the development of the Reference CDMS will be directed towards the needs of developing countries.

Once the Reference CDMS has developed sufficient functionality, effort will be required to support the transition of other CDMS and their data into the Reference CDMS.

While the ideal goal may be to reduce the number of CDMS from around 96 down to one (the Reference CDMS), it should be recognised that:

- Commercial organisations may believe that their investment in their Proprietary CDMS is viable and continue development. This is to be encouraged and the organisations invited to participate within the Reference CDMS community to help fund, develop and re-use Reference-CDMS components.
- In the future it is anticipated that several variants of Open-CDMS may evolve to support specific deployment goals, e.g. Developing Countries; Regional Climate Centres; or Global Climate Centres. Careful coordination will be required to ensure that all variants continue to use and extend the same set of source components in order to maximise the re-use of investment.

Recommendation:

- This work will require seed funding to start with, together with the maintenance of a core CDMS team to lead the project as the open source community is being established. This funding requirement is also briefly discussed in the Data Rescue and CDMS Resource Plan. Additional funding will be required as address the long term sustainability of the Reference CDMS.

Intervention is required

Given the problems in global implementation of CDMS outlined above in the section entitled '*What is the current global experience of CDMS implementation*', together with the level of investment required for consistent global CDMS capability and consistent data, and the potential number of CDMS requiring this investment, it is considered necessary for WMO to intervene and provide leadership in CDMS development.

The ET-CDMS Policy Statement on CDMS noted the level of CDMS development required and recommended for convergence on one open source CDMS called Open-CDMS.

Therefore a key aspect of this CDMS Implementation Strategy is to provide for the rationalisation of 96 different CDMS to a number that is more affordable to maintain for the long term.

In short this CDMS Rationalisation Strategy allows for:

- The maintenance of the three most widely used open source CDMS: CliDE, MCH and Climsoft for several years;
- The development of consistent CDMS and climate data functionality that is conformant with WMO #1131 within the Reference Open Source CDMS, Open-CDMS. Priority will be given to the needs of developing countries;
- The convergence of at least CliDE, MCH and Climsoft capability into Open-CDMS;
- Proprietary and other organisations to be invited and participate with the development of Open-CDMS within an open source community; and
- Proprietary organisations to continue their development of their closed source CDMS, guided by Open-CDMS development, should they believe their investment justified.

Coordination of CDMS Rationalisation

It will take a concerted effort over many years to guide the rationalisation of available CDMS to ideally a single Reference CDMS.

The following is recommended to help coordinate and guide the CDMS Rationalisation:

CDMS Secretariat

A secretariat will be required to support CDMS Rationalisation activites. This is expected to include:

- Running, moderating and supporting the Open Source CDMS Community;
- Acting as the initial Governance body for Reference CDMS development until the CDMS Community matures. When appropriate, this role will transition into the Open Source CDMS Community;
- Overseeing CDMS Development, coordinating activities as required;
- Guiding the transition of legacy CDMS to the Reference CDMS;
- Organising and running CDMS Developer meetings; and
- Providing a core CDMS Development team.

CDMS Developer meetings

Regular CDMS Developer meetings are recommended to coordinate activities resulting in evolution of the Reference CDMS:

- The primary focus of these meetings will be to coordinate the development of the Reference CDMS, currently named Open-CDMS; and
- There will also be a need to coordinate activities between current open source and other CDMS, to ease their transition and migration to Open-CDMS.

Recommendations

It is recommended that:

- Annual CDMS Developer Meetings are held for at least five years. However at times of high activity where close coordination is required, several additional meetings may be required; and
- CDMS Developer Meetings allow participation by open source and proprietary CDMS; NMHS and other open source developers to allow issues to be adequately investigated and coordination actions agreed.

Acting Nationally

From the CDMS Strategy Vision 2018-2023:

> *Achieve a step-change in the management of climate data by:*
> *Acting nationally to implement and evolve consistent and sustainable CDMS that will address national requirements for climate data, and in addition help to address regional and global needs.*

Current situation

As outlined above, the 2012 CDMS Survey has shown that there are significant issues with CDMS implementations globally, with the most pressing issues affecting developing countries.

These issues have resulted in decades of indifference, with an implicit policy that expects all countries to develop and maintain their own CDMS software. For many developing countries, this expectation is impractical.

The 2012 CDMS Survey, the 2017 DARE/CDMS Resource Plan and follow up work by ET-CDMS have shown that many developing countries have more than one CDMS implemented, and in use. This situation has often arisen through aid projects of short duration that have implemented a CDMS with inadequate planning, training and follow up support. Historical climate data is often not migrated as part of the CDMS implementation project.

Consequently, once the project team has left and there is no support available on how to use and maintain the new CDMS, climate data management staff often fall back on what they know and understand. This leaves historical climate data spread across multiple tools and is typically very difficult to integrate for long-term time-series.

Coordinated CDMS Deployment is required – globally

While it is reasonable to expect developed countries to fund their own CDMS implementation and ongoing support and training costs, it is time for a rethink on how CDMS are deployed and supported, particularly within developing countries.

Noting the Lessons Learned discussed in the 2017 DARE/CDMS Resource Plan and the Australian Government experiences in implementing CDMS within the Pacific region, economies of scale can be achieved by coordinating CDMS deployment to developing countries across a region. For example:

- Costs for planning, travel for multiple return visits, training, development, support, etc. can be spread across a number of countries, making the CDMS implementation costs cheaper on a per country basis; and
- Regional communities of practice can evolve and be nurtured to facilitate a situation where regional partners help each other.

Therefore this CDMS Implementation strategy proposes a coordinated global CDMS Deployment. The main aspects *'Acting Nationally'* for this strategy include:

- CDMS Communications Plan;
- National CDMS implementations
- Coordination of CDMS Implementation;

CDMS Communications Plan

A CDMS Communications plan is required to ensure that a coordinated message is delivered to all NMHS, GCC, and RCC explaining why work is required to upgrade their CDMS; what work is expected and in what timeframe. There will also need to be follow-up communications to coordinate and monitor CDMS implementation activities.

It is important that these communications activities provide NMHS with sufficient time to plan for and to budget for their CDMS activities. Many NMHS may require at least two years advanced warning.

Therefore this is a priority task in the CDMS Implementation Strategy.

Recommendation:

It is recommended that the CDMS maturity model and self-assessment tool discussed in the 'Think Globally' section of this strategy be established as a priority to support this communications plan and subsequent activities.

National CDMS implementations

It is expected that each NMHS will need to undertake activities to implement a modern CDMS that is compliant with WMO #1131.

Typically, countries will be expected to fund their own CDMS related implementation and maintenance activities through their internal IT practices.

Issues to consider for a CDMS Implementation at a single country are summarised in WMO #1131 CDMS Specification, Section 10, Considerations pp. 134-139.

However, for a sustainable CDMS Implementation, the work does not stop there. **The CDMS must be supported, regularly maintained and upgraded to retain its benefit to the organisation**. This is a key factor in ensuring that a CDMS Implementation is sustainable.

Therefore it is recommended that:

- NMHS ensure that they maintain an internal capability to support and maintain their CDMS Implementation. Regular updates to CDMS functionality are expected, particularly during the WMO Community's transition to modern CDMS that are compliant to WMO #1131.

It is also recommended that:

- In specific circumstances it may be appropriate to allocate funding to support qualifying Developing Countries (DC) in their CDMS implementation and maintenance activities for a period of five years to enable the DC to build their own internal CDMS capability. This may perhaps be via the RCC responsible for supporting that DC;
- Appropriate guidelines for applying for and approving requests for such support for assistance are clear and unambiguous;
- Data Rescue activities are included as part of the CDMS implementation to assist the organization to capture their historical records;

- The support is conditional, requiring the recipient of the assistance to commit to provide access to their data in accordance with Resolution 60 (Cg17); and

- If a DC does not have a CDMS implemented currently, one of the three current open source CDMS (CliDE, MCH, or Climsoft) be implemented initially and later upgraded when the modern open source CDMS Open-CDMS becomes available. This will provide DC some initial CDMS capability, while waiting for a modern CDMS to become available.

NB: Modern CDMS compliant with WMO #1131 are estimated to be available within approximately three years of this strategy being approved, provided that funds are found to support the 'Think Globally' aspects of this strategy'.

Coordination of CDMS Implementation

The goal of the coordination of CDMS implementation activity is to achieve a **step change** in climate data management globally by providing a supportive environment and facilitating the gradual transition to next generation CDMS over at least a five year timeframe.

This activity will need to ensure that regular CDMS related metrics are maintained to demonstrate that the effort and investment is achieving the desired goals. This is discussed further in the 'Think Globally' aspect of this strategy, under the section titled '*Establish clear, unambiguous metrics to measure improvements*'.

It is recommended that CDMS Implementation activities are coordinated to ensure widespread implementation of modern, sustainable CDMS and to build and nurture regional communities of practice that are able to help each other. It will also create an environment where **partnering** between NMHS could flourish:

- It is proposed that regional activities are coordinated via Regional Climate Centres, with the support of Global Climate Centres.

- It is proposed that global activities are coordinated by WMO via GFCS processes, also with the support of Global Climate Centres; and

- As part of this coordination process, it may be appropriate to consider what level of specialised support arrangement that an RCC or GCC can provide on behalf of peers within its region. This may help to simplify CDMS Implementation, particularly for DC;

- It is anticipated that one or more CDMS Flying Squads may be required to provide support to NMHS and RCCs in their CDMS planning and possibly implementation activities. Such a team would probably have experts with a broad range of technical skills required to support a CDMS implementation. WMO #1131 CDMS Specification, Section 10, Considerations p135 provides example skills likely to be required.

It is further recommended that the overall coordination of this activity may be appropriate activity for GFCS.

Funding

where does the funding come from

- potential funding sources
    - GFCS
    - CSIS

- Green Climate Fund
- Voluntary contribution to Data Rescue and CDMS Fund as per Resource Plan
- other?

Appendix A

to WMO Strategy to Improve CDMS Implementation

Architectural overview of climate data

Appendix A: What is a Climate Data?

Definition

Climate data allows a scientist to understand for a given location[4], both:

- The meteorological conditions related to a specific event; and
- The climatic conditions that have occurred, or may occur, over an extended period of time.

Conceptually

Conceptually, climate data may comprise all of the following:

1. A time-series of observations of a specific phenomenon, from a single sensor, including:
    - Estimations of the value of an observed property;
    - Calculations derived from the estimations of value;
    - A detailed understanding of the conditions under which the observations were made; and
    - Any changes that have been made to the observations (e.g. due to Quality Assurance processes).
    - The phenomena being observed (or estimated) may be one of many variables of interest to the climate science community, for example the GCOS Essential Climate Variables[5].
2. A collection of the time-series observations described at #1, of the same phenomenon or calculation, at the same time steps, perhaps in the form of a set of discrete coverages (as a time-series).
3. A time-series representing the distribution of values of the collection of time-series observations represented as perhaps a:
    - Continuous two dimensional coverage;
    - Continuous coverages as three dimensional cubes;
    - Continuous coverage as n-dimensional models;
    - An ensemble, comprising a number of models; or
    - The outputs of some analytical process as a time-series/coverage/cube/model.

Using the description of climate data above, when a peer reviewed paper is published, a scientist needs to be able to refer reviewers to:

- The analytical data, which underpin the paper (perhaps the n-dimensional continuous coverage time-series at #3);
- The quality assured observations data that the continuous coverages at #3 were derived from at #2 and #1 (with details as to why each change to the Quality Assured observations described at #1 were made);

[4] The term location is used here in a very broad sense. A location could be a site, local area, country, region, continent, or the whole world.

[5] https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables

- The ‘raw’ observations data described at #1, with details as to the conditions, sensors, etc., that the observations were made under; and
- The details and version of the software source code that was used to manipulate the data at #1, #2, and #3.

This is not a trivial data management problem to address, however its resolution will provide a solid data management foundation for future climate science.

Appendix B

to WMO Strategy to Improve CDMS Implementation

CDMS Rationalisation

Appendix B: CDMS Rationalisation

Noting:

- The findings of the 2012 CDMS Survey that there are:
    - Significant issues with the implementation of CDMS globally; and
    - Estimated to be 96 CDMS for 190 countries and that these CDMS provide different and inconsistent capabilities.
- ET-CDMS work from early 2017 that shows that 42 developing countries are reliant on open source CDMS that are under severe financial constraints, have single points of failure in key developers, and are consequently at risk of failure;
- Recommendation 11.4 of WMO #1131, CDMS Specifications pp 148-150. This recommendation was for WMO to sponsor and support open source approaches to CDMS development.
- WMO #1131 was approved by Cg-17.
- ET-CDMS Policy Statement on CDMS dated November 2015[6], approved by CCl Management Group.
- The ten year CLIMAT analysis 2007-2016 that indicates significant issues with climate data management globally.

Background to CDMS Rationalisation

There are estimated to be 96 different CDMS in use globally. These CDMS have been developed in the absence of CDMS Specifications and it is understood that none have the functionality expected in a modern CDMS as defined in WMO #1131.

It is understood that all CMDS developers are facing severe constraints in obtaining sufficient funds to develop and maintain their systems.

It will be a waste of scarce funds and resources to attempt to develop the same capability within each of the 96 CDMS in order to upgrade them with the functionality that is expected within a modern CDMS.

It is time for a rethink on how CDMS are developed.

This CDMS Strategy proposes that NMHS and other interested parties **collaborate and pool funds** to develop and maintain a single Reference CDMS that is designed to provide the functionality expected within a modern CDMS as defined by WMO #1131.

This Reference CDMS is nominally called Open-CDMS as at the date of this Strategy document.

This Reference CDMS would be developed and maintained using tried and tested open source software practices.

For an overview of why an open source approach has been selected for this CDMS Implementation Strategy, see WMO #1131, Recommendation 11.4, pp148-150.

[6] http://wis.wmo.int/file=2253

This approach will reduce the long term financial outlay that WMO, NMHS, and other organisations expend to develop and maintain various CDMS.

WMO #1131 is designed as a set of inter-related components that can be combined to provide the capability expected within a modern CDMS. Many of these components describe functionality that is often readily available within existing open source and proprietary software packages.

Therefore it is not expected that the Reference CDMS will be a single monolithic software application, but a collection of software applications that are developed, extended, integrated and bundled to provide the required CDMS functionality.

This approach will allow for future innovation within software components, allowing individual components to be upgraded and replaced, rather than replacing the complete CDMS.

Similarly this approach will allow the Reference CDMS to be scaled to meet a range of needs, from the needs:

- Of developing countries; to
- The advanced technology needs of many NMHS, Regional Climate Centres and Global Climate Centres.

The work in designing, developing, integrating and upgrading the Reference CDMS will need to be funded and coordinated.

This CDMS Strategy expects that this Reference CDMS work will be undertaken within an open source software community and that any software development is undertaken as an open source software project. This approach will:

- Establish and provide the rules and governance on how CDMS work is undertaken;
- Allow collaboration by any interested party, including: UN; WMO; NMHS; proprietary and open-source commercial organisations; open source software developers and communities; and members of acadaemia.
- Spread the cost and burden of developing the Reference CDMS across many organisations;
- Ensure that the results of any effort is not locked up in closed source, black-box proprietary solutions and that the Intellectual Property and results of any investment can be re-used in the future with little or no constraints;
- Ensure that the software developed to manage, analyse and provide climate data is open for review and scrutiny;
- Facilitate a transfer of knowledge for those interested in learning more about CDMS development; and
- Develop a robust and collaborative Open Source Software community to facilitate the long-term viability and sustainability of the Reference CDMS.

The Reference CDMS is nominally called Open-CDMS. This name may be reviewed in the future.

Priority in the development of the Reference CDMS will be directed towards the needs of developing countries.

Once the Reference CDMS has developed sufficient functionality, effort will be required to support the transition of other CDMS and their data into the Reference CDMS.

While the ideal goal may be to reduce the number of CDMS from around 96 down to one (the Reference CDMS), it should be recognised that:

- Commercial organisations may believe that their investment in their Proprietary CDMS is viable and continue development. This is to be encouraged and the organisations invited to participate within the Reference CDMS community to help fund, develop and re-use Reference-CDMS components.
- In the future it is anticipated that several variants of Open-CDMS may evolve to support specific deployment goals, e.g. Developing Countries; Regional Climate Centres; or Global Climate Centres. Careful coordination will be required to ensure that all variants continue to use and extend the same set of source components in order to maximise the re-use of investment.

Recommendation:

- This work will require seed funding to start with, together with the maintenance of a core CDMS team to lead the project as the open source community is being established. This funding requirement is also briefly discussed in the Data Rescue and CDMS Resource Plan. Additional funding will be required as address the long term sustainability of the Reference CDMS.

Intervention is required

Given the problems in global implementation of CDMS outlined above in the section entitled '*What is the current global experience of CDMS implementation*', together with the level of investment required for consistent global CDMS capability and consistent data, and the potential number of CDMS requiring this investment, it is considered necessary for WMO to intervene and provide leadership in CDMS development.

The ET-CDMS Policy Statement on CDMS noted the level of CDMS development required and recommended for convergence on one open source CDMS called Open-CDMS.

Therefore a key aspect of this CDMS Implementation Strategy is to provide for the rationalisation of 96 different CDMS to a number that is more affordable to maintain for the long term.

In short this CDMS Rationalisation Strategy allows for:

- The maintenance of the three most widely used open source CDMS: CliDE, MCH and Climsoft for several years;
- The development of consistent CDMS and climate data functionality that is conformant with WMO #1131 within the Reference Open Source CDMS, Open-CDMS. Priority will be given to the needs of developing countries;
- The convergence of at least CliDE, MCH and Climsoft capability into Open-CDMS;
- Proprietary and other organisations to be invited and participate with the development of Open-CDMS within an open source community; and
- Proprietary organisations to continue their development of their closed source CDMS, guided by Open-CDMS development, should they believe their investment justified.

Benefits to the WMO Community from this CDMS Intervention

There are many benefits that will be realised by the WMO Community by undertaking this CDMS intervention. These include:

- *Financial*
    - In the long term, the costs to develop and maintain CDMS will be minimised, with the cost burden shared by many organisations. There will be no need to develop and maintain 96 separate CDMS.
- *Agility*
    - By having a responsive CDMS Community and a coordinated and sustainable CDMS implementation framework it will be much easier to support future WMO initiatives:
        - The Reference CDMS software can be modified to support the new requirements;
        - The CDMS changes can then be deployed during a scheduled upgrade cycle; and
        - Providing a mechanism to speed up adoption of new WMO policies considerably.
- *Consistent Climate Data*
    - Modern CDMS will have the tools and functionality embedded to:
        - generate standard climate data products that use the same semantic data definitions and approved peer reviewed algorithms; and
        - find and exchange data in accordance with WIS-2, WIGOS and other related WMO requirements.
- *Strategic*
    - This intervention will lay the foundations for a future CSIS / WIS-2+ framework that will support globally federated and consistent climate data services similar to those outlined in the CDMS 10+ year vision above.

Benefits to NMHS from this CDMS Intervention

There are many benefits that will be realised by individual NMHS by undertaking this CDMS intervention. These include:

- *Financial*
    - NMHS can treat the CDMS as an off the shelf software product, but without costs for annual software license renewals.
    - The NMHS does not have the burden of developing and maintaining CDMS software and support for new WMO initiatives by themselves. They can share the burden with their peers within the open source CDMS community.
- *Agility*
    - Should an NMHS have a specific need and priority, they can work within the CDMS Community to find peers with a similar need and share the costs of the development and maintenance.

Benefits to Climate Science from this CDMS Intervention

There are also many benefits that will be realised by the Climate Science Community by undertaking this CDMS intervention. These include:

- *Consistent Climate Data*
  - Consistent climate data is available from all NMHS that implement CDMS that are conformant with WMO #1131;
  - This results in less time spent of getting climate data into a consistent form to support climate analysis and more time available to undertake the actual analysis; and
  - Concepts related to data provenance are built into the CDMS, making it easier to validate the data used within an analysis.

Coordination of CDMS Rationalisation

It will take a concerted effort over many years to guide the rationalisation of available CDMS to ideally a single Reference CDMS.

The following is recommended to help coordinate and guide the CDMS Rationalisation:

CDMS Secretariat

A secretariat will be required to support CDMS Rationalisation activites. This is expected to include:

- Running, moderating and supporting the Open Source CDMS Community;
- Acting as the initial Governance body for Reference CDMS development until the CDMS Community matures. When appropriate, this role will transition into the Open Source CDMS Community;
- Overseeing CDMS Development, coordinating activities as required;
- Guiding the transition of legacy CDMS to the Reference CDMS;
- Organising and running CDMS Developer meetings; and
- Providing a core CDMS Development team.

CDMS Developer meetings

Regular CDMS Developer meetings are recommended to coordinate activities resulting in evolution of the Reference CDMS:

- The primary focus of these meetings will be to coordinate the development of the Reference CDMS, currently named Open-CDMS; and
- There will also be a need to coordinate activities between current open source and other CDMS, to ease their transition and migration to Open-CDMS.

Recommendations

It is recommended that:

- Annual CDMS Developer Meetings are held for at least five years. However at times of high activity where close coordination is required, several additional meetings may be required; and
- CDMS Developer Meetings allow participation by open source and proprietary CDMS; NMHS and other open source developers to allow issues to be adequately investigated and coordination actions agreed.

Suggested CDMS Rationalisation Goals

Short Term Goals

The short-term CDMS Rationalisation goals for year one to two, are suggested as:

- Establish a process to ensure funding and maintenance of existing open source CDMS to mitigate against risk of failure in CDMS that are used by many developing countries. These CDMS are CliDE, ClimSoft and MCH. As part of this process it may be appropriate to:
    - Fund one to two key people for two to three years in each of the three open source CDMS to ensure that key knowledge on the CDMS is maintained to facilitate transition to the Reference CDMS; and
    - Ensure WMO representation on the relevant Project Steering Committees to represent WMO's interest in the open source CDMS.
- Prepare for Open-CDMS in year one. Tasks will need to include:
    - Establish initial governance processes for managing Open-CDMS. These processes will eventually transition to the CDMS Open Source Community when it matures;
    - Establish the Open-CDMS contributor's agreement that assigns all intellectual property that is developed in Open-CDMS to an appropriate entity, together with functionality to manage these agreements;
    - Establish an initial Open-CDMS web site;
    - Establish a 'strawman' roadmap for Open-CDMS for discussion at the intitial CDMS Developers Meeting;
    - Prepare an Open-CDMS options paper that will provide recommended Open-CDMS development and funding models; and
    - Establish initial seed funding to allow Open-CDMS development to commence development with a core team, while longer term funding arrangements are investigated and implemented.
- Establish and launch a CDMS communications plan to ensure that CDMS issues and the need for change is well understood;
- Prepare and run the initial CDMS Developer meeting to discuss issues and agree on priority tasks to meet the CDMS needs of developing countries;
- Provide feedback on the development of relevant WMO initiatives, issues and priorities. Also provide WMO with feedback on experiences when trying to implement support for these initiatives; and
- Commence work on the Reference CDMS, currently named Open-CDMS prototype development activities.

Medium Term Goals

- Continue maintenance of existing open source CDMS to mitigate against risk of failure in CDMS that are used by many developing countries. These CDMS are CliDE, ClimSoft and MCH;
- Continue CDMS Developer meetings to guide the development of the Reference CDMS, Open-CDMS;
- Continue support for the Open-CDMS open source community;

- Phase out support for maintenance of open-source CDMS: CliDE, ClimSoft and MCH;
- Continue work on Reference CDMS development and maintenance activities;
- Guide transition of CDMS to Open-CDMS; and
- Provide feedback on the development of relevant WMO initiatives, issues and priorities. Also provide WMO with feedback on experiences when trying to implement support for these initiatives.

Long term Goals

- Continue CDMS Developer meetings to guide the development of the Reference CDMS, Open-CDMS.
- Continue work on Reference CDMS development and maintenance activities.
- Guide transition of CDMS to Open-CDMS;
- End support for maintenance of open-source CDMS: CliDE, ClimSoft and MCH.
- Transition responsibility for Open-CDMS maintenance to the Open-CDMS open source community.
- Reduce from 96 different CDMS towards perhaps just one CDMS.