World Meteorological Organization

COMMISSION FOR CLIMATOLOGY

Expert Meeting on WMO Climate Data Management System Strategy
Geneva, Switzerland, 12-14.XII.2018

EM-WCDMSS/Doc N(3)

Submitted by:
Steve Foreman

14.XII.2018

Strategy for the reference CDMS tool set

This document will become Document 2(1) of the 2019 workshop.

Review comments

Version/ No	Reviewer	Issue	Notes	Response
0.1/1	Stuber	Para 4.2.1. We could let the door open not only for “franchise” CDMS (Climsoft, MCH, CliDE), but also for proprietary CDMS (CLIWARE, CLISYS, CLIDATA, CLDB, etc.) that may consider to join the OPEN CDMS project. Or even some **regional organizations** as (Agrhymet, Asecna, SADC) that use proprietary systems and would like to move to an open source. In that case we should also include those proprietary CDMS into possible migration.		Amended para 4.2.4 and inserted new para before it.
0.1/2	ET-DRM	Emphasize the project principle of “free open source” development		Amended para 4.1.1.2. Also updated the “Development Governance” document to include this in the purpose statement.

DISCLAIMER

Regulation 43

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

Regulation 44

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

Chairperson, Publications Board

World Meteorological Organization (WMO)

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

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

CH-1211 Geneva 2, Switzerland E-mail: Publications@wmo.int

Strategy for the reference CDMS tool set

1. CONTEXT FOR THE STRATEGY

1.1 DECISIONS BY WMO CONGRESS AND OTHER BODIES

1.1.1 Resolution 16 (Cg-XVI) *Climate data requirements* stated:

Decides that priority will be given to:

…….

(6) Producing and compiling World Weather Records and WMO climate normals based on improved methodologies and guidelines to assist Members in their computation and dissemination;

(7) Modernizing climate data management and related services, including archiving, discovery, access and retrieval as part of the evolving implementation of the WMO Information System and ensuring interoperability amongst various programme data, as well as with sector application data;

(8) Developing human and technological capabilities of developing and least developed countries to fulfil their mandate in climate data collection, management and exchange;

Requests Secretary General to promote collaboration and partnership with international agencies and programmes and mobilize extra-budgetary resources for capacity-building activities relevant to climate data modernization including DARE and CDMS projects …

1.1.2 Resolution 60 (Cg-17) WMO policy for the international exchange of climate data and products to support the implementation of the Global Framework for Climate Services stated:

Urges Members:

(1) To provide the additional GFCS relevant data and products that are required to support and sustain the operational climate services as the core element of the Framework and WMO initiatives at the global, regional and national levels and, further, as mutually agreed, to assist other Members to enhance access to GFCS relevant data and products and in the provision of climate services in their countries; such additional GFCS relevant data and products are listed in the annex to the present resolution and could have conditions associated with their use, such as restrictions on their use for commercial purposes, attribution of their source or licensing;

(2) To establish funding mechanisms, including new investments, for sustaining the network of stations and sensors needed for the global observing systems for climate, and also the maintenance and operation of the data preparation and management systems necessary to support the implementation of the present resolution.

1.1.3 The Sendai framework for Disaster Risk Reduction (WMO, 2015) includes:

Promoting the collection, analysis, management and use of relevant data and practical information in line with national circumstances and making use of space and in situ information that results from maintained and strengthened in situ and remotely-sensed Earth and climate observations;

Ensuring dissemination of reliable data in an appropriate format and accessibility of non-sensitive information, taking into account the needs of different categories of users (including social and cultural requirements, in particular, gender);

Promoting the further development of and investment in effective, nationally compatible, regional multi-hazard early warning mechanisms, where relevant, contributing to GFCS, and facilitate the sharing and exchange of information across all countries;

Supporting relevant UN entities to strengthen and implement global mechanisms on hydrometeorological issues in order to raise awareness and improve the understanding of water-related disaster risks;

Promoting international cooperation for DRR and enhanced coordination of respective strategies of UN entities and other international and regional organizations, especially in developing countries, in particular, the least developed countries, small island developing states, landlocked developing countries and African countries.

1.2. WMO COMMISSION FOR CLIMATOLOGY

1.2.1 At its seventeenth session [CCl-17], the Commission for Climatology considered the approaches being taken to climate data management, and in Resolution 4 (CCl-17) agreed that the Manual on High Quality Global Data Management Framework for Climate should be finalized. The same resolution endorsed the Climate Data Management System strategy that is reproduced in Annex 1.

1.2.2 The strategy acknowledged that almost 50% of WMO Members faced operational issues with the Climate Data Management Systems (CDMS), almost 25% of WMO Members did not use an electronic database to manage climate data, and that there were almost 100 different CDMS in use by WMO Members.

1.2.3 CCl recognized that continuing to maintain and develop the current diversity of CDMS was, when looked at globally, more expensive than it need be and wasteful of scarce skills and resources. CCl accepted that developing a single reference open source CDMS would allow Members to pool resources, reducing duplication of effort.

1.2.4 CCl reiterated that climate data now encompasses observations of all the Essential Climate Variables and also standard WMO Products, gridded data, outputs from numerical models (for example reanalyses, climate predictions and projections), and space-based and impact data. Recognizing that, in this document the term “climate data” includes all types of data recognized as such by CCl-17.

1.3. GLOBAL FRAMEWORK FOR CLIMATE SERVICES

1.3.1 As part of the implementation plan for the Global Framework for Climate Services (GFCS) [WMO 2018b] there is a project to implement a Climate Services Toolkit (the project description is reproduced in Annex 2). One of the components of that project is climate data management, and the plan recognizes CDMS instances such as ClimSoft and CliSys as part of their tool kit.

1.3.2 GFCS, therefore, recognizes a requirement for a common tool set of climate data management tools as essential to the provision of climate services.

1.4. WMO INFORMATION MANAGEMENT

1.4.1 WMO Information System 2.0 Strategy [WMO, 2017] recognized the limitations of the first implementation of the WMO Information System (WIS), and introduced additional functionality to simplify access to shared information. In WIS 2.0, services will continue to be provided by centres operated by WMO Members coordinated through the Global Information System Centres (GISCs), but increased use of web services and cloud technologies will mean that the services they coordinate will become simpler to use for non-scheduled exchange of information. Put simply, WIS 2.0 will make it easier for users to obtain access to the data they need when they need it and, especially if the volume of data is very large or the calculations complex, will provide interfaces for users to use the computing facilities that are co-located with the data, thus overcoming the problems of transferring large volumes of data over inadequate telecommunications links.

1.4.2 WIS 2.0 will define interfaces that are generic, flexible and based on standards that are widely used in information technology. The very flexibility of those standards will pose a challenge to users of climate information. With so many options to choose from, working out which combination will work for each data provider would be a major task. Providing standard configurations of how those interfaces should be used for climate data would greatly simplify the tasks of data providers and data users.

1.5. CLIMATE DATA MANAGEMENT SYSTEM SPECIFICATIONS

1.5.1 In 2014 WMO published the Climate Data Management System Specifications ([WMO-No. 1131]. These provide a common framework of good practice for managing climate data that includes setting policies, defining procedures and acquiring the tools needed to support these. Even though [WMO 2014a] does not address all the activities needed for managing climate data, no current CDMS application fully implements it.

2. OBJECTIVES OF THE STRATEGY

2.1 This strategy aims to ensure that all nations have access to software tools they need to help them manage climate information to internationally accepted standards of stewardship in order to safeguard the climate record upon which those nations can build climate services.

2.2 In particular, the strategy seeks to provide a reference CDMS tool set that assists all WMO Members, particularly those with limited access to ICT skills and funding. In this context, the phrase "reference CDMS tool set" refers to a suite of software supporting climate data management that might be a single computer program, but is more likely to consist of separate modules that perform separate tasks and that work closely together. The tool set would be designed to:

2.2.1 Comply with WMO and other international standards for managing climate data;

2.2.2 Comply with WMO and other international standards for recording metadata associated with climate data;

2.2.3 Use internationally agreed specifications where these are available for processing the data;

2.2.4 Implement controls on adding, modifying or deleting data;

2.2.5 Maintain an audit trail of all actions that impact on the data in the system;

2.2.6 Demonstrate that the climate data and the products generated from them have been managed in line with good practice;

2.2.7 Implement internationally recognized good practice for information security;

2.2.8 Report metrics specified by WMO that contribute to the WMO maturity index for management of climate data.

2.3 In addition, the strategy seeks to avoid unnecessary costs in developing and operating climate data management tools through cooperative development, maintenance and support. It also seeks to manage financial, operational and sustainability risks associated with bespoke or proprietary software.

3. REFERENCE CLIMATE DATA MANAGEMENT TOOL SET

3.1 OPEN SOURCE

3.1.1 In 2011, in response to the impacts of a theft of emails from the University of East Anglia, the UK Parliamentary Committee on Science and Technology recommended "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, 2011]. The principle of open review of science underpinned this conclusion, and a key point made by the committee was that review of the computer codes that underpin climate science is as important as review of the algorithms themselves. The source code for the reference CDMS toolkit should, therefore, be available for review by anyone.

3.1.2 The intention of the UK Parliament was to promote confidence in the conclusions drawn by climate scientists. Inevitably, there will be bugs in the software of the reference CDMS tool set. Making the source code available for anyone to review will expose these. The impact of these bugs on how the reference CDMS tool set is viewed, and, more importantly, on how conclusions based on data stored and processed by the tool set are trusted would depend to a great extent on how those developing and using the tool set respond to reports of bugs.

3.1.3 Publishing the source code at the same time as a scientific paper in which it was used was published could reduce the confidence in the validity of the paper itself if bugs were found in the code. The functionality provided by the reference CDMS tool set is, however, not in itself a key differentiator between different teams of scientists or new methods of analysing the data. It could be thought of as more like a compiler for a computer language – necessary for performing the science, but of little interest in itself to other than specialists in that technology.

3.1.4 An alternative approach to publishing the source code of the reference CDMS tool set is to make the evolving code available for scrutiny while it is being developed and correcting reported bugs as part of the development process. This cannot guarantee that the final version of the code is bug-free, but it does increase the chances of bugs being found and corrected. Any remaining bugs could still, of course, call the validity of the conclusions of papers using data handled by that component of the reference CDMS tool set, but would be seen clearly as errors rather than as attempts to bias the results.

3.1.5 Open publication of source code is problematic for software vendors that rely on sales of software licences, although there are many examples where vendors make some of their code openly available, while keeping the code for "premium features" private. The "premium features" business model relies on income from the additional functionality to cover the costs of developing the openly available portions of the software.

3.1.6 At the heart of the reference CDMS tool set is the requirement to implement standard functionality and algorithms and to exchange information using standard formats and interfaces. There would be little to distinguish between the offerings of different vendors for that functionality – differentiation would take place through the provision of additional functionality, ease of use, and integration with other business processes. This is very similar to the situation in the market of operating systems in the early 2000s; vendors of computers based on the Unix operating system were finding it increasingly difficult to justify the costs of maintaining their own flavour of the operating system, and resolved this by contributing to the development of the linux operating system, sharing resources with competitors but reducing their own cost base for what had become an undifferentiated component of their offering.

3.1.7 Developing the reference CDMS tool set as an open source project, while requiring changes to the traditional approach of WMO and National Hydrological and Meteorological Services (NMHSs) to system procurement and development, would:

3.1.7.1 Make the source code for the tool set openly available;

3.1.7.2 Potentially open the pool of developers beyond NMHSs;

3.1.7.3 Allow vendors to reduce their cost base and concentrate on adding value for their customers;

3.1.7.4 Reduce the challenges facing users of CDMS in maintaining their systems against hardware and software obsolescence.

3.1.8 Many of the advantages of developing the reference CDMS tool set as an open source project will be lost if the tool set relies on proprietary components. Where possible, it should use open source tools, for example its database management system (where PostgreSQL and MySQL community edition are candidates). Where use of proprietary tools is unavoidable, the tool set will need to take care to make migration to an alternative as straightforward as possible.

3.2 MODULAR

3.2.1 The reference CDMS tool set is intended to be just that, a tool set. Although it has to be released in a form that allows users to run a fully functioning CDMS "out of the box", the components of the tool set also have to be available for use in other contexts. It should, for example, be possible to develop, test and validate an improved algorithm for data homogenisation using the reference CDMS tool set to provide the supporting interfaces and functionality. Among other advantages, this would allow proposed amendments to WMO regulations, standards or recommended practices to be tested in the environment actually used by WMO Members in advance of formal endorsement of those amendments – and allow Members to implement the amendments within a short time of their endorsement.

3.2.2 Creating the "out of the box" reference CDMS tool set will take time – probably several years. Some components will be ready much earlier than others (for example, the data storage component will need to be available early in the development, even if only in a limited form). Early release of these components would benefit the user community, who could build them into their working methods.

3.2.3 A modular architecture for the reference CDMS tool set, with a well-defined and strongly-governed set of interfaces for each modular component, would allow parallel development of the components of the tool set, users (and vendors) to implement additional functionality beyond that of the tool set, and WMO to test the impact of proposed changes in climate data management practices.

3.3 AUDITABLE AND AUDITED

3.3.1 Aside from the expectation of overall reductions in the cost of developing, maintaining and operating a CDMS, the reference CDMS tool set is intended to increase the confidence in the outputs of climate science and climate services. The development process for the reference CDMS tool set has to support confidence in the tool set itself, and the method of operation of the tool set has to support confidence in the data held and processed by it.

3.3.2 To build confidence in the reference CDMS tool set itself, the development process should be open, with decisions and the reasons for them clearly visible, the design and code of the tool set clearly documented, openly reviewed with the review comments and responses to them visible, open reporting of bugs and the responses to them, and a decision making process that prioritises integrity and international best practice.

3.3.3 Although the process in the preceding paragraph is open, the importance of the reference CDMS tool set is high enough that independent audit of the development process is justified. The purpose of that audit would be to confirm to stakeholders that the open development processes of the project were actually being used effectively and to identify areas in which the processes should be improved.

3.3.4 The reference CDMS tool set has to include internal logging and tools to analyse the logs that can be used to audit the provenance of information managed by the tool set.

3.3.5 In addition to supporting validation of how data management by the system has been handled, the tool set must also support reporting on the effectiveness of the data management system itself by reporting standard statistics on the types of operation that have been performed, the volume and type of data held and other metrics that the user community consider important. The metrics should include those in the WMO maturity matrix for climate data management. [A link to maturity matrix is needed – though the full WMO matrix has not yet been approved - the NCEI stewardship model is at http://dx.doi.org/10.2481/dsj.14-049]. Metrics should be available through a standard interface to assist WMO in gathering summary statistics, and to demonstrate that the centre is handling data in an appropriate way.

3.4 LIFE-CYCLE MANAGEMENT

3.4.1 Implementing new versions of software in an operational system is often complex, costly and time consuming – even if the new version differs only slightly from the previous one. Those WMO Members most in need of the reference CDMS tool set are generally those least able to manage the change process to use a new version of the software. Nevertheless, it will be essential to introduce new versions of the tool set to implement fixes for bugs and security weaknesses, to introduce new functionality, and to update the algorithms in line with changes in good practice or WMO regulations. This implies a need for strong version control and tools to assist users with upgrading their systems and automated deployment of updates to the CDMS tool set. Automated deployment needs an extremely effective testing regime to be in place to manage the risks to end users of deploying faulty updates. Users choosing to make modifications to the software and not have them included in the core version would need to be responsible for their own software updating regime as there could be no guarantee that central updates would work with local changes. Experience with CliDE has shown that the risks and difficulties of updating modified software outweighed any benefits of local modifications to the software.

3.4.2 If successful the reference CDMS tool set will be used as a resource by other software systems, whether developed in-house by users (with the risks noted above for automated software updates), by the open source community or by vendors. This places a constraint on the development of the tool set itself – as far as possible it should be backwards-compatible, so that any of these additional systems that were developed for an earlier release of the tool

set will continue to work correctly when a new release of the tool set is implemented. Over time, this constraint will become increasingly difficult to meet, so it is likely that there will be an infrequent requirement for "breaking releases". In this case, it may be necessary to continue to maintain previous releases for bug fixes and security issues.

3.4.3 A feature of open source software is that it can be modified by its users. This is one of its great strengths. That freedom causes a potential issue for the reference CDMS tool set. One of the aims of the tool set is to give the broader community confidence in the validity of the software being used to manage climate data. If the user of the software is able to modify it, what does that mean for the confidence of the community? One solution [Fogel, 2018] is to apply a trade mark to the "official" version of the software that is issued by the open source project. Although users can still modify the software any way they wish, and distribute it to others, they must not call it by the same name as the "official" version. This approach has been taken by the OpenWIS® open source project, for example. The control over the naming of the "official" version also allows the project to enforce its own version identification system.

3.4.5 WMO regulations, standards, specifications and recommended practices will be embodied in the reference CDMS tool set. Open review will provide one mechanism to confirm that the software complies with them. There is no guarantee, however, that an open source review will examine that compliance specifically. Clear documentation of how the compliance is implemented would assist WMO in confirming that the software did, if used correctly, comply with its requirements. WMO might choose to audit that compliance and publish its findings as a way of increasing confidence in the tool set and results obtained using it.

4. DEVELOPMENT OF THE REFERENCE CLIMATE DATA MANAGEMENT TOOL SET

4.1 OPEN SOURCE PROJECT

4.1.1 Implementing the reference CDMS tool set as open source software involved several types of consideration:

4.1.1.1 Ownership of the software – even though the software would be issued under an open source licence, that licence has to be granted by a legal entity, and that entity has to be granted licences from each of the contributors for the components of the system they provide. Even with its privileges and immunities, WMO might not be in the best legal position to own the software – to overcome some potential legal issues the openWIS® association chose to set up a legal entity separate from any of its proponents in a jurisdiction whose laws were sympathetic to the concerns of developers of open source software;

4.1.1.2 Licence for the software – all open source software should have an explicit licence that states what users may do with the software. The choice of licence can strongly influence whether developers contribute to the project, or whether users are willing to implement the software. Some open software is licensed so that users may change, distribute and modify the software with no restrictions whatsoever. Other licences allow any changes to be made and distributed but that place restrictions on the type of licence the resulting system can have. The licence for the reference CDMS tool set should have a licence that does not limit the ability of the user to modify the software ("free" open source). The GPL 3 licence, for example, requires the software and any modifications to it to be made available with the same licence, but it explicitly allows that software to be bundled alongside other software that may have a different licence. The design aim of enabling interaction with the reference CDMS tool set through well-defined interfaces would be compatible with that type of licence.

4.1.1.3 Long term sustainability – users of the software will rely on it for many decades, and they need to have confidence that there will be an active community of developers supporting it;

4.1.1.4 Speed of development – any project that makes slow progress is unlikely to attract investors. The investors in open source projects are the developers, and if they do not see the fruits of their labour they are likely to move to other projects. So, not only do the users benefit from rapid progress with the project, but the project stands more chance of success if it can deliver rapidly, especially in its early stages.

4.2 EVOLUTION NOT REVOLUTION

4.2.1 "But first, look around" is the advice given by Fogel (2018) on how to start a new open source project. By this he means that if there is another open source project that is trying to achieve the same outcome as a project under consideration, there needs to be a very good reason for starting a new project rather than contributing to an existing one. There are already three CDMS systems with open licences (the "franchise" systems): CliDE, ClimSoft and MCH. Each of these is struggling to obtain the resources they need to be sustainable, but none of them has a clear lead over the others in terms of functionality or compliance with WMO-No. 1131. Each of them has a user base that is in urgent need of updating its CDMS.

4.2.2 Regardless of how the reference CDMS tool set is developed, users of the franchise systems will at some time need to convert to the tool set.

4.2.3 Rather than choosing one of the franchise systems as the starting point for the reference CDMS tool set, an alternative would be to develop an overall architecture for the tool set and the key interfaces between its modules. With that architecture in mind, the franchise systems would then be studied and their components mapped onto the architecture. These components would then be available for the project to adapt or take as inspiration. An early part of the reference CDMS tool set project should be to allow migration of dataset content between the databases of the franchise systems and the tool set. Together with an interface between the tool set's database and the data access routines of the franchise systems, this would allow a smooth migration path between the franchise systems and the tool set.

4.2.4 ET-CDMS found that 45% of WMO Members had created their own CDMS, and that others were using proprietary CDMS such as CLIWARE, CLISYS, CLIDATA and CLDB. The organizations developing and maintaining these systems might also want to contribute to the development of the reference CDMS tool set.

4.2.5 With a smooth migration from the franchise systems and other contributing systems to the reference CDMS tool set available to them, climate data managers should be reassured that they can implement any of the franchise systems as a first step towards implementing the tool set, allowing them to use an electronic database to help them manage their data.

4.3 ENCOURAGING CONTRIBUTIONS

4.3.1 A major challenge for ensuring that the reference CDMS tool set is sustainable is attracting developers to contribute to it.

4.3.2 Most open source projects need to complete the early stages of development before becoming a truly open source project (Fogel, 2018). Developers need to see that the project is serious in its intent to deliver. With a convincing start, other developers would be more likely to join the project so that it becomes self-sustaining.

4.3.3 Potential sources for the initial development include:

4.3.3.1 NMHSs

4.3.3.2 Members of HMEI

4.3.3.3 Universities or research institutes

4.3.3.4 Contracts with software developers funded through donors.

4.3.4 An *ad hoc* working group of WMO experts would need to see the project through to its formal launch as an open source project.

4.3.5 As a sustainable open source project, the majority of contributors to the ongoing project should self-nominate and self-fund. There will, however, be an ongoing core of activities that are necessary for WMO to meet its ambitions for the tool set but for which progress is not being made. By establishing a Trust Fund, WMO could pay for developer time to implement such functionality if required.

4.3.6 One of the objectives of the reference CDMS tool kit is to give users confidence that the software is handling climate data according to good practice. A key element of achieving this is to have good documentation, but Fogel (2018) points out that the typical developer of open source software in not motivated to write documentation, and that if they do so, the documentation often makes the assumption that the knowledge level of the reader will be higher than that of most actual readers. Fogel also notes that those in the best position to write useful documentation are those who have only just learned how to use the software. This ties in closely with one of the other objectives of WMO that is only indirectly related to the tool set – improving training and development of users in climate data management. Building writing and reviewing documentation of the tool set into the training activities supporting climate data management would have the dual benefits of delivering documentation written from the user perspective, and improving the understanding of those being trained in the use of the tool set as well as increasing their confidence in their ability to use it.

4.4 MAKING DECISIONS

4.4.1 Decision making on the project

4.4.1.1 Like other projects, open source projects need decision making processes. Unlike other projects, there are usually few formal ties between those contributing to the project and the project itself – often a developer is a member of the project because she wants to contribute, not because of a contractual commitment to do so. This means that successful open source projects tend to use consultative decision making processes in which those with the most successful record of making contributions to the project tend to have the greatest influence in the final decision – a "meritocracy". This can be uncomfortable for many organizations, such as WMO, that are accustomed to more formal hierarchies in decision making, but history shows that open source projects tend to be more sustainable if they use an inclusive decision making process (Fogel, 2018). Although it may be necessary for individuals to make rapid decisions in the early days of the project, the individuals making those decisions have to be respected by the development community, and the project should move towards decisions by consensus as far as possible.

4.4.2 Decision making about what the project is to support.

4.4.2.1 Although WMO-No. 1131 contains guidance on good practice for climate data management, it is not exhaustive. It is in the nature of software development that, as functionality is translated into computer code, practices that were thought to be well-described turn out to have vagaries, and new needs for recommended practice are identified. Even the fast-track procedures available in the Manual on WIGOS are not fast enough to allow formal

WMO decision making on the time scales needed by an open source development process. Joint decision making between the project and representatives of the WMO decision making procedures will be necessary to confirm that proposals are "good enough" to be implemented in the system as "candidate practice", with the expectation that they may be modified as they are tested in actual use, and both the "candidate practice" and the software might be modified as the outcome of that learning is applied as a WMO practice.

4.5 SEPARATION OF DATA WITHIN THE REFERENCE CDMS TOOL SET

4.5.1 All organizations deploying the reference CDMS tool set will need to train their IT staff, data managers and data users in how to use the tool set to support the organizations' processes and assess the impacts of changes in their working practices. Using their operational CDMS for this introduces a significant risk to the integrity of the data within the system. This can be mitigated by careful design of the system to allow a single instance of the software to support multiple logical instances that use distinct databases within a single database management system, distinct logging systems and distinct configuration controls. Users of one logical instance would not be able to access data or configuration information in a different logical instance other than by using an explicit export/import mechanism.

4.5.2 In addition to providing users with functionality they need, such a separation of data and configuration would allow a single instance of the software to serve multiple organizations, making it suitable for deployment as an "application as a service" in which the reference CDMS tool set could be made available using cloud technologies in a way that meant that it appeared to data managers and users that they were using their own dedicated system.

4.6 CREATING A SUPPORTIVE ENVIRONMENT

4.6.1 Implementing the reference CDMS tool set will be a significant undertaking. In addition to the obvious need for people with the right skills to write and test the software and to help Members implement it, developing the tool set will also highlight where existing guidance does not provide enough detail for it to be implemented in software.

4.6.2 Developing the reference CDMS tool set as a open source project is not "free". Software developers and experts in climate data management will need to spend time contributing to the project. Many, if not most, of the developers and experts will come from organizations that already have software supporting their climate data management activities – and the expertise of the developers is likely to be highest in those components of a CDMS that have already been implemented within the organization. Expertise in other components is likely to be found elsewhere. So, the needs of the project for rapid development of robust software are most likely to be met if organizations contribute staff time to implement those aspects of a CDMS that the organization already uses (although, of course, there may be some aspects of the tool set that have not been implemented by any of the organizations willing to contribute to its development). This creates a paradox for the organization – they will be asked to invest in developing components of the software that duplicate what they already have. The paradox is resolved because their investment will be used as "payment" to receive the fruits of the investments made by other organizations for things that those organizations already have. This logic differs greatly from that normally used when making investment decisions, so the project (and the developers supporting it) will need help in educating other developers and organizations in the benefits of investing in the reference CDMS tool set rather than continuing to evolve their own solutions.

4.6.3 Software developers think like software developers. Users think like users. Unless users of the software are involved closely with the development throughout the design, coding and testing phases, the software is likely to be difficult to use or, worse, actually hinder users in performing their tasks – and even if the software actually does what is needed, it may not be

documented in a way that users can understand. Users and their managers will need to be involved in the software development process.

4.6.4 In the absence of guidance (such as the detailed calculation of a climate normal), the details of how software developers implement the generic guidance will be driven by their own experience. Although the review procedures for the tool set would provide some scrutiny of how the guidance had been implemented, the fact that the review would be of computer source code would make it difficult for those who are most expert in how to process climate data to take an active part in the review. As a result, the methods of calculation used in the tool set, though themselves acting as a reference standard, may not reflect a consensus on how climatologists would like the calculations to be performed. This can only be avoided by close interaction between the software developers and experts in handling climate data capable of making authoritative recommendations on the short time scales needed by the software project.

4.6.5 Taking the above into account, the project will need help in communicating with all stakeholders, including standard-setting bodies, heads of organizations, resource managers in organizations, climate data experts, software developers and people involved in the day-to-day activities associated with climate data management, if it is to obtain the resources it needs and the standards that it can implement.

5. OUTLINE PLAN FOR THE REFERENCE CLIMATE DATA MANAGEMENT TOOL SET

5.0.1 The following pragraphs outline in bullet form what should be included in the implementation plan for the reference CDMS tool set.

5.1 COMMUNICATIONS

5.1.1 The communications plan has to address the needs of the following stakeholders.

5.1.1.1 **WMO Technical Commissions and** Secretariat – what CDMS tool set is, regulatory support needed, expertise required. Progress, achievements and priority needs (relevant to TCs etc).

5.1.1.2 **PRs** – benefits of CDMS tool set; why they should invest in the tool set; why they should implement the tool set (and how); benefits of supporting use of tool set by others; progress and achievements and priority needs.

5.1.1.3 **Climate service providers** – benefits of CDMS; how the tool set supports climate service delivery; how to contribute and influence the project; progress, achievements and priority needs (relevant to climate service providers).

5.1.1.4 **Climate data providers** – benefits of CDMS tool set; how CDMS tool set helps better use of their data; need for metadata; how simplified data policies can increase uptake of their data; how CDMS tool set can simplify management of their data; progress, achievements and priority needs (relevant to climate data providers)

5.1.1.5 **Climate data managers** – benefits of CDMS tool set; how they can contribute to CDMS tool set; how to use CDMS tool set; how CDMS tool set can include their local procedures; progress, achievements and priority needs (relevant to climate data managers)

5.1.1.6 **Software developers** – aims of CDMS tool set; how to take part in project; status of project; priorities for development, items for review, bug reports (and

reporting), discussion of issues being addressed by project; progress, achievements (both development and impacts of implementing) and priority needs;

5.1.1.7 **Trainers** – what CDMS tool set is about; needs for training in the use of the tool set; using the tool set to support training activities; contributing to the development and documentation of the tool set;

5.1.1.8 **Funding agencies** – benefits of the CDMS tool set; including implementation of the CDMS tool set as part of infrastructure projects; building maintenance into funding projects (pay now to fix current issues – others will pay in the future to fix future issues); achievements and impacts of funding maintenance.

5.2 MANAGEMENT AND RESOURCES

5.2.1 The following topics related to management and resources should be addressed in the implementation plan.

5.2.1.1 Entity to "own" the software and the project (formal management by representatives of key agencies, such as HMEI, WMO, IOC, the developers, …)

5.2.1.2 Initial and ongoing management structure for the project

5.2.1.3 Managing a trust fund

5.2.1.6 Funding/people for initial development

5.2.1.6.1 Contributing staff

5.2.1.6.2 Contracts for development – individuals or companies

5.2.1.6.3 Individual volunteers

5.2.1.7 Software environment to run the project (configuration management systems, communication systems, wiki, bug tracking, ….)

5.2.1.8 Agree software licence (eg GPL 3) and Intellectual Property arrangements – specialist legal advice is needed.

5.2.1.9 Decision making for making guidance specific ("*de facto standards*")

5.3 BUILDING THE CRITICAL CORE

5.3.1 The "critical core" of the reference CDMS tool set is the mimimum set ot tools needed to support compliance with the CDMS specification (WMO-No. 1131).

5.3.1.1 Run by WG of experts

5.3.1.2 Architecture, interface definitions

5.3.1.3 Database, ingest, output data

5.3.1.4 Recommended quality control (climate extrema, time continuity, ….….)

5.3.1.5 Simple statistics

5.3.1.6 Logging of transactions

5.3.1.7 Information security (access controls, backup and retrieval, integrity protection, ...)

5.3.1.8 Build, manage and maintain a software code respository…

5.3.1.9 Interfaces to allow use with climate statistical packages (such as Climpact).

5.3.1.10 Mechanism for agreeing algorithms and processes to be encoded in the reference CDMS tool set. (This could be one of the objectives of the proposed CDMS Community of Practice)

5.4 MAINTAINING CURRENT CAPABILITY

5.4.1 The project developiong the CDMS tool set has to provide a migration path for users of the franchise CDMS

5.5 MAINTAINING THE REFERENCE CDMS TOOL SET

5.5.1 The reference CDMS tool set has to be maintained so that it remains current against changing regulations, standards, recommended practices, IT security concerns, IT hardware and software, and other external changes.

5.5.1.1 Management and funding of ongoing maintenance of the reference CDMS tool set (this is confined to the "officially" released trunk version of the software; and modifications or additions made by users would not be supported by the project).

5.5.1.2 Support to users of the reference CDMS tool set (perhaps through the Regional Climate Centres).

5.5.1.3 Contribution mechanisms for WMO Members, other organizations (public or private) and individuals to assure long term maintenance of the tool set.

5.6 INCLUDING THE CDMS SPECIFICATIONS

5.6.1 The reference CDMS tool set is intended to support compliance with the CDMS specifications (WMO-No. 1131). The implementation plan has to address the following topics.

5.6.1.1 Analyses of gaps between CDMS specifications and the reference CDMS, and between the specifications and current practice as reflected in the maturity matrix.

5.6.1.2 Monitor project development and use the Trust Fund to implement functionality to reduce the gap that is not being filled "organically".

5.7 ADDING NATIONAL RESPONSIBILITIES

5.7.1 The addition of support for national responsibilities to the reference CDMS tool set is about managing expandability and software substitutions in a controlled way – APIs, alternative versions of modules/algorithms, additional metadata (which may not be disclose-able outside the organization), additional data types...

5.8 SUPPORT FOR IMPLEMENTATION

5.8.1 The reference CDMS tool set has to assist users in implementing the system in their own organizations. Some of this will be technical (for example, using standard installation tools to

install the software), but some will have to take the form of advice and assistance with the implementation.

5.8.1.1 Is support for implementation part of the open source project, or a separate capacity development activity?

5.8.1.2 How can developers be encouraged to help with support – or is it developers that are needed? (perhaps a peer-peer network of climate data mangers would be better, or some Regional Climate Centres might accept the responsibility)

5.8.1.3 Installing and maintaining a software system needs skilled ICT staff and secure, resilient ICT equipment that may not be available to all potential users of the reference CDMS tool set. It would technically be possible for an organization to offer access to a cloud-based implementation so that users could control and manage their own data but not have to manage the software and hardware themselves. Such an approach would increase the number of users who could benefit from the tool set.

BIBLIOGRAPHY

WMO, 2014a: Climate Data Management System Specifications. Geneva, Switzerland. World Meteorological Organization. WMO-No. 1131.

WMO, 2014b: Annex to the Implementation Plan of the Global Framework for Climate Services – Climate Services Information System Component. Geneva, Switzerland. World Meteorological Organization. 62pp.

WMO, 2015: Towards Substantially Reduced Disaster Risk in 2030. Bulletin Vol 64 (2) – 2015. World Meteorological Organization.

WMO, 2017: WMO Information System 2.0 strategy. World Meteorological Organization. WMO-No. 1213.

WMO, 2018: Commission for Climatology. Abridged Final Report of the Seventeenth Session. Geneva. 10-13 April 2018. WMO, Geneva, Switzerland. World Meteorological Organization. WMO-No. 1216.

Fogel, K., 2018: Producing Open Source Software. How to Run a Successful Free Software Project. (Accessed 5 November 2018).

UK Parliament, 2011: The Reviews into the University of East Anglia's Climatic Research Unit's E-mails. London, UK. UK Parliament. .

ANNEX 1: CLIMATE DATA MANAGEMENT STRATEGY ENDORSED BY CCL-17

The following text is reproduced from the report of the seventeenth session of the Commission for Climatology.

Annex to Resolution 4 (CCl-17)
Climate data modernization

1 Climate Data Management System Strategy (Concept)

1.1 Summary of gap analysis

A CCl Expert Team on Climate Data Management Systems (ET-CDMS) survey revealed that WMO Members operate almost 100 different Climate Data Management Systems (CDMSs), developed in the absence of consistent WMO CDMS specifications. Based on this survey it is estimated that almost 50% of WMO Members face operational issues with their CDMS and almost 25% of Members do not use an electronic database to manage climate data.

1.2 Vision

The vision for Climate Data Management Systems (CDMSs) is to achieve a step-change in addressing capability required to respond to many questions of societal needs for climate information and services; and acting nationally to implement and evolve interoperable and sustainable CDMSs that will address national requirements for climate data, and in addition help in addressing regional and global needs.

1.3 Concept of the Strategy

(a) Establishment of a sustained global framework for Climate Data Management including recommended practices and standards as well as CDMS related activities

In view of the importance of well-managed climate data for almost all climate applications and services, ET-CDMS developed Climate Data Management Specifications (WMO-No. 1131) in alignment with WMO relevant technical regulations and the Guide to Climatological Practices (WMO-No. 100). Based on these, a framework is required that comprises a base collection of technologies, policies and institutional arrangements that facilitate the management of, and access to, consistent climate data, thereby establishing clear governance arrangements around CDMS activities; consistent community agreed formal data and process definitions; and metrics to measure CDMS improvements. Strengthened coordination to develop CDMS as an integral part of WMO overall information management will greatly facilitate the framework's evolution.

(b) CDMS rationalization

It is considered very expensive and wasteful of resources to maintain and further develop a multitude of CDMSs worldwide, and to establish compliance to the above CDMS specifications. This element includes a revision of the current ways to develop and maintain CDMSs by establishing a coordinated close collaboration of partners involved including pooling resources to develop and maintain a single Reference open-source CDMS based on tested software practices.

(c) *Coordination of CDMS deployment*

While CDMS implementation and maintenance in general constitutes a national responsibility of the NMHS, it is recognized as a basic element of capacity development for a number of developing countries. Coordination of CDMS deployment at global and regional levels will greatly improve efficiency by providing a supportive environment and facilitating the gradual transition to next-generation CDMSs.

(d) Development of a CDMS communications plan

The communication plan should ensure that a coordinated message is delivered to all NMHSs, international data centres and partners, including the GFCS Partners Advisory Committee, explaining why work is required to upgrade CDMSs, and what work is expected over what time frame for delivering timely and high-quality data for international programmes; in particular the World Climate Programme and GFCS.

ANNEX 2: GFCS CLIMATE SERVICES TOOLKIT

The implementation plan for the Climate Services Information System of the Global Framework for Climate Services (GFCS) [GFCS-CSIS] includes a project to define, build and make available a Climate Services Toolkit to all countries. The text describing that project is reproduced below.

Project 2: Define, build and make available a Climate Services Toolkit to all countries

Activity:

This project aims to identify, collect, enhance and package a high-quality set of knowledge products, software tools, public domain datasets and related training materials to assist d latest scientific and technological advances, as needed by users and stakeholders, and will promote consistency and quality in the products and services developed through CSIS. Many institutions will contribute to the toolkit, with considerable effort required to develop, test and complete the materials for widespread use. WMO, through CCl, will coordinate the compilation, production and distribution of a Climate Services Toolkit. This activity reflects relevant priorities under the DRR (risk analysis and assessment as well as early warning) and Health Exemplars (indeed, as an overarching, multi-sectorial requirement).

Objectives:

To ensure that climate-sensitive sectors in any country have access to the most up-todate,

reliable and consistent climate information and products that meet their basicneeds at least;

To provide a conduit for technology transfer to developing countries, enabling their access to the latest methods, techniques and information required for CSIS activities and products;

To increase the effectiveness, consistency and quality of CSIS activities and outputs;

To identify, collect, enhance and package a high-quality set of knowledge products, software tools, and related training materials, i.e., a climate services toolkit;

To distribute the toolkit to CSIS entities and advise them on its application;

To establish a procedure for maintaining and updating the toolkit (as users increase their participation in the GFCS, and increasingly benefit from climate information, their requirements will likely evolve, which may require developing new tools to meet these requirements. Moreover, as research advances are made the toolkit must be updated to accommodate new materials).

Benefits:

Implementation of GFCS at national scales will impose considerable demands on the service providers, including NMHSs. Having a toolkit based on standards and good practices to support CSIS activities will improve efficiency and raise capacity of service providers, and will ensure that the information and products developed for and provided to users is reliable,consistent (through time and across regions) and of high quality. A toolkit can be kept up to date with new tools, information and methods, and therefore will enable all CSIS providers to take advantage of research advances. The datasets included in the toolkit will enable more countries to develop their national products and should encourage improved data sharing. The availability of the toolkit, with training materials, should reduce the need for expensive capacity building. The Climate Services Toolkit will also make training workshops more focused,tangible and efficient in imparting the operational skills.

Deliverables:

A toolkit, consisting of knowledge products; bespoke software for data management,data analysis (including indices), climate monitoring, prediction, downscaling and verification, with the requisite training materials; a set of standards, and a certification process for new tools;

A collection of standard public domain datasets (e.g. global gridded data, monthly SST ata, etc.), as well as data generated by data rescue, digitization homogenization and CDMS projects for inclusion in the toolkit;

A plan for maintenance and updating of the toolkit and its datasets.

Current activities:

CCl has an Expert Team on CSIS that is developing the toolkit as one of its key deliverables. Other CCl teams are developing software (e.g. CCl ET CCDI for climate indices, and ET CRSCI for sector-specific indices). WMO Members and research and academic institutionshave developed, inter alia, ClimSoft and CliSys for data management; CMT for monitoring,CPT, PRECIS and SCOPIC for forecasting, downscaling and verification and so on.

Indicators and Assessment Measures:

Number of countries with access to and using the toolkit;

Number of training workshops based on Climate Services Toolkit;

Number of operational CSIS products using the Climate Services Toolkit;

Number of contributors to Climate Services Toolkit.

Participants:

This work will be conducted by CCl and CBS experts, representatives from advanced NMHSs,academic and research institutions, on CSIS, RMP and relevant O&M aspects.