WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

A PROPOSED LONG TERM RESOURCE PLAN FOR CLIMATE DATA MANAGEMENT AND RESCUE ACTIVITIES

WMO OMM

World Meteorological Organization
Organisation météorologique mondiale

Table of Contents

Note: *This document was developed and written by Dr William James Wright (lead author) from the Bureau of Meteorology, Australia, co-chair of the Commission for Climatology (CCl) Open Panel on Climate Data Management. Helpful information and feedback Rob Allan and ACRE colleagues, Bruce Bannerman, , Simon McGree and Celine Becker from the Australian Bureau of Meteorology. The document was endorsed by Tom Peterson and Barbara Tapia (President and vice-President CCl.*
The work on Data Rescue and Climate Data Management Resource Plan is coordinated by the WMO Data Management Applications Division.

Executive summary

In response to WMO's Executive Council's decision at its sixty-eighth session, June 2016 (Decision 39 (EC-68)) this paper presents a proposed resource plan for the sustainable support of climate data management systems (CDMS) and data rescue activities (the Plan). The initial focus in this Plan is on undertaking or supporting climate data rescue and CDMS implementation over the next five years, in developing and least-developed countries (D & LDC) that have recently expressed their need for assistance with securing and managing their data. Most of these countries are in Regions I, V and the Indian Ocean Region. Preliminary indicative costings per country over five years to implement these remedial actions have been developed, informed by actual and estimated costs for similar work carried out in Regions I and V. In addition, a number of engagement, collaboration and capacity-building actions that complement the financial investment and lay the groundwork for longer-term sustainability are also described.

The proposed approaches will build on existing technologies and practices, of which there are many fine examples already in place. **Appendix A** reviews some of these current initiatives around CDMS and Data rescue. **Appendix B** summarises some of the lessons learned and recommendations from undertaking CDMS and data rescue projects in the past; these provide some background to several of the assumptions made in the costings.

For CDMS, noting the emphasis on D & LDCs, it is proposed that the three current Open Source CDMSs ClimSoft, MCH and CliDE be the primary basis of the deployment, as these are already widely deployed in many D & LDCs, and were purposefully designed and built with the special needs of D & LDCs in mind. For data rescue, the proposal is to build on the needs and collaboration mechanisms identified by WMO data rescue activities (including regional initiatives such as INDARE and WACA-DARE), global programs such as ACRE and IEDRO, and individual country-initiated projects such as the Australian Government- funded data rescue efforts in the Pacific. It is also proposed to strengthen, maintain and populate the existing I-DARE portal (www.idare-portal.org), which provides a global picture of data rescue needs, and to extend this to accommodate details of data rescue needs and projects in the hydrological and marine domains as well. For the sake of efficiency it will be important to collaborate with evolving development work in data rescue and data management (an example being the EU's Copernicus initiatives), as well as being mindful of the potential advantages offered by emerging digital technologies.

Based on costings for existing initiatives around CDMS deployment and data rescue in the Pacific, supplemented by information from similar initiatives in Africa, an indicative cost of approximately **CHF 349K per country** over five years is indicated for in-country activities. The overall cost for the in-country program *including 35 countries constituting a benchmark for estimating the costs* is estimated at just over **CHF 12 Million**. This sum is expected to cover CDMS implementation and training (including follow-up training), maintenance, bug-fixes, and some development work, and to initiate data rescue activities including securing hard-copy records under accepted "good practice" archival conditions, as well as initiating and providing some funding for inventorying, digitisation[1] and imaging programs. An important caveat is that the actual needs and amount of work needed will vary substantially across countries, depending on what has been achieved so far,

[1] In the context of this paper, "digitisation" refers to the transcription of data from hard copy forms into a data management system via key-entry.

identified needs, and the quantities of data needing to be rescued. Additional cost of just over **CHF4.2 Million** is estimated for upgrading and maintaining I-DARE, and for development work on OpenCDMS (refer below). Details of, and assumptions made, in formulating these costings are provided in **Appendix C**. It is important to re-emphasise, however, that successful implementation of the Plan is not just about financial support, but ensuring that the countries' Governments and their NMHSs are suitably engaged, and that collaboration mechanisms and capacity development activities required are set in place. Details of these additional requirements are briefly outlined below, and elaborated on in **Appendix D**.

Looking to the longer term beyond five years, the measures identified in Appendix D are germane to establishing an ongoing sustainable infrastructure for environmental data and its management. An ongoing, reliable source of funding for especially CDMS maintenance will also be required: it is not practical at this stage to specify an amount, although it is expected to be substantially less than the initial investment. One of the measures proposed in Appendix D is to invest in converging CDMS technologies around an Open Source system that is fully compliant with the CDMS specifications (WMO no 1131), but is suitably designed to accommodate data from other environmental domains such as hydrology. Accordingly, the Plan proposes investment to develop a prototype for OpenCDMS during the five years, although subsequent development of the full system should be funded independently of this initiative. The longer-term strategy would be to support existing CDMS in developing countries via other elements of the Plan, while gradually transitioning into implementing and supporting OpenCDMS.

Finally, the costings and projections quoted here specify only the recommended requirements, not where the funding comes from. Such sources may, however, include international development agencies, and donations including through the Global Framework for Climate Services (GFCS), I-DARE etc.

Purpose

To outline a proposed approach to the development of a sustainable, cost effective resourcing strategy for Climate Data Management Systems (CDMS) and data recue activities, with initial emphasis on activities in D & LDCs that have expressed urgent needs to WMO to secure, digitise, manage and archive their climate data over the next five years. The paper also proposes recognising and working towards a collaborative approach for data rescue and management with other technical Commissions, notably hydrology and in the area of marine data, given the commonality of the challenges and work required between atmosphere, ocean and water.

Background

For climate data to effectively underpin climate services, both at the individual country and global level, it must be securely managed and accessible and fit for purpose. For this reason, WMO's Commission for Climatology and other Technical Commissions have in recent years placed considerable emphasis on the rescue, secure archival, digitisation and management of climate data, and over this period there has been considerable progress made in data rescue projects and the development and implementation of CDMSs. An overview of such initiatives is provided in **Appendix A.** Despite these efforts, much data remains locked up in hard copy forms, often at considerable risk of loss or damage, while many countries either do not have, or have difficulty supporting, effective, secure CDMSs. This latter reflects that the skills and resources required to sustainably maintain systems

are often lacking in developing countries, which places at risk the ongoing secure archival of their climate data holdings.

For these reasons, Decision 39 (EC-68) endorsed the development of a long-term resource plan to address the need for cost-effective and sustainable resourcing for climate data management activities in support of the GFCS, specifically the implementation and maintenance of modern CDMSs, and the recovery and digitisation of historical climate records. Such a strategy will require a variety of approaches, not just financial support, but will require investment in setting up the arrangements and practices required.

The challenges faced and work required for securing and managing climate data[2] apply not only to atmospheric data, but to other domains, notably hydrological and marine data, especially in the area of data rescue. Therefore, to ensure an efficient and cost effective approach to data rescue and management approaches, it is further proposed to work collaboratively with representatives from the relevant Commissions to identify potential ways of converging the approaches and technologies toward data rescue and management. For example, the existing I-DARE data rescue portal could readily be extended to include details of hydrological data rescue needs and projects. In the longer term existing CDMS, which mostly focus on the management of atmospheric data, could be expanded in scope to include marine and hydrological data (thereby becoming a more generic Environmental Data Management System, EDMS).

Benefits of the Plan

The benefits of investment in the rescue and management of climate data are numerous, and include the following, among others:

(1) Access to more historical data, in discoverable, accessible and exchangeable climate datasets, will better improve the ability of the global climate community to understand and monitor climate variability and change, including supporting UNFCCC processes ;
(2) Improved availability of historical and real time datasets, especially at sub-daily, daily and monthly timescales, will provide better information on extremes, a fundamental input for risk assessment in support of national disaster risk mitigation;
(3) Improved data for reanalysis and climate modelling helps improve the understanding of past climate and its variations, and supports more realistic projections of future climate change;
(4) More data managed more effectively at the national level improves the ability of Member countries to provide efficient, effective and scientifically sound climate services to their people on an ongoing basis;
(5) It is expected that modern CDMSs operated at national level, coupled with integrated data management infrastructure at global and regional levels, will represent an efficient approach to managing, sharing and using data and information regionally and globally.

These benefits have been long-known, and as Appendix A demonstrates, have inspired many excellent initiatives in the development and implementation of CDMS and in data rescue. The problem is that such efforts often tend to be project-based, with funding only available for the defined life of the project (typically just a few years). This means that,

[2] In the context of this paper, "climate" data will normally be interpreted as referring to the atmospheric domain, but can be interpreted as applying broadly to any data domain with a time series, i.e. historical, component.

especially in D & LDCs, data recovery efforts are not sustained, and in terms of CDMS, countries tend to revert to pre-existing arrangements such as spreadsheets, or cease managing their data. Sustainable resourcing, as opposed to *ad hoc* funding for individual projects (with its inevitably temporary funding), will address the gaps outlined below. At the same time, many countries still don't currently have an effective CDMS or data rescue project in place.

Specific actions to address gaps

The following is a summary of the opportunities that an effective long-term resourcing plan would enable:

Data Rescue: The need to sustain (and in many cases initiate) data rescue efforts such as securing at-risk data, timely technology migration, creating and publicising inventories, and digitisation and imaging. In D & LDCs, the Plan would support the following major steps forward:

a. taking prompt initial action to secure vulnerable hard-copy records against loss or damage, through proper archival processes;
b. providing funding and training to support personnel involved in carrying out or facilitating data-rescue efforts, including imaging and digitisation;
c. helping develop an ongoing within-NMHS and regional capacity to maintain data rescue activities, including through community of practice and capacity-building activities;
d. better population of the I-DARE portal with information about data to be rescued, as a means of identifying gaps and establishing priorities.

CDMS: Implementing a CDMS in a D & LDC is not simply a "set and forget" effort, but requires ongoing support to ensure the required maintenance, administration, software upgrades, and staff competencies are maintained. In particular, such support should enable:

a. access to suitably-trained CDMS administration staff within or outside the NMHS;
b. ability to apply upgrades, patches, bug-fixes etc.;
c. migration from existing systems to the new CDMS;
d. customisation and maintenance of CDMS implementation in line with each NMHS's IT environment, which will enable more efficient(and therefore sustainable) integration with existing workflows and practices, and better alignment with emerging climate service user needs (this is why the costings provided in Appendix C assume that initial in-country training is required, with follow-up in-country training after two years);
e. regular dedicated regional workshops to encourage ongoing training, development of regional communities of practice, etc.;
f. for the purposes of this analysis, because the most urgent need for CDMS deployments are in D & LDCs, it is assumed that Open Source systems already widely deployed in such countries will form the basis of future deployments (e.g., ClimSoft in RA I; MCH in RA III and IV; CliDE in RA V). If systems based on proprietary technology such as Oracle are deployed, additional licensing costs will apply. Nevertheless this should not deviate from the over-all average trend in the costing provided in this plan since these systems, as a general rule, are considered to be supported as part of existing bilateral, multi-lateral, or regional collaboration

frameworks where WMO would play a facilitation role as part of this Resource Plan (see section : "*other activities required to ensure sustainability*"

Implementation Plan

The implementation of this Plan seeks foremost to extend existing best-practice in the areas of data rescue and CDMS. For the first five years, the Plan will focus mainly on implementing CDMS and data rescue projects in countries that have informed WMO that they face particularly high risks of data loss. Indicative costings for these implementation activities, and their maintenance over a five year period, are provided under "Costings" below. The main elements may be summarised as follows:

A. An initial in-country assessment visit for each country;

B. An initial in-country implementation visit to install a CDMS, provide initial training, and to instigate data rescue activities by securing hard-copy documents, and initiating the development of inventories, document ordering, digitisation and imaging (as described in Appendix B, this enables the training of staff in their own work environment, and the training of multiple staff);

C. Initial funding to hire contract staff for imaging and digitisation activities (noting that in many cases the allocation here will not be sufficient to complete all digitisation and imaging);

D. A follow-up in-country visit for software upgrades, refresher training etc;

E. Allocation for a certain number of "contingency" trips each year to address urgent problems or issues;

F. Annual regional workshops to provide further training, and maintain and enhance collaboration and a community of practice;

G. Replacement of the PC on which the CDMS sits after a standard five year period;

H. In addition, funding is required to maintain, and expand, the I-DARE portal, and to develop a prototype of OpenCDMS.

It should be noted that the above elements represent a full implementation cycle for CDMS-data rescue activities, and in practice many countries already have some of the elements in place (e.g., a CDMS installed; initial assessment study completed; digitisation or imaging projects underway).

At the same time, a number of other measures are required, which include promotional, communication, collaboration and ongoing capacity-building activities. These will complement the above activities, but are also essential to support longer-term sustainability. They apply variously at the national, regional and global level, and are described under "*Other activities required to enhance sustainability*" below.

Also to support long-term sustainability, access to a reliable pool of funding (similar to the concept of the VCP) will be required, especially to support CDMS maintenance, administration, interventions when needed, and the maintenance of I-DARE. To ensure the continuity of such a fund, it is recommended that funding from several different sources be combined into an ongoing pool. It is not practicable at this stage to provide estimates of how much ongoing funding will be required, as this would be heavily influenced by factors such as technical advances (e.g., the possibility that improved Optical Character

Recognition software could expedite digitisation), and the degree to which some of the other initiatives such as OpenCDMS can be implemented.

Costings

Appendix C provides a first order estimate of costs required to initiate CDMS implementation and data rescue activities, and maintain them for five years for countries identified as high priority. The costings here are based on actual CDMS deployment and data rescue activities in mostly Small Island Developing States within Region V, with some additions and modifications based on experience in Region I. While noting that overheads applied within the NMHS implementing the project within RA V may be higher than elsewhere, in general we believe these provide a sound basis for estimating costs in other D & LDCs.

The costings span a five year period, which is the estimated time before the PC on which the CDMS is installed within each country would need to be replaced. They cover installation of a CDMS and configuration to each country's IT and working environment, along with training, maintenance, refresher training and some development work for the CDMS. For data rescue, the costings cover: initially securing hard-copy records under "good practice" archival conditions, and initiating and contributing some funding for, inventorying, digitisation and imaging programs using mainly externally- sourced hired help such as retirees and students.

The estimated cost per country over a five year period is estimated at 349K. In practice, the costs would be lower in many countries where, for instance, a CDMS is already in place or overheads are lower. However it is important to recognise that, for data recue in particular, the costs of imaging and digitising records will in some countries be much greater than the allowance here, unless significant NMHS staff resources were available to undertake the work.

Other activities required to enhance sustainability

Whilst access to funding to implement, and/or maintain CDMS and data rescue projects is vital, this in itself is not sufficient. For long-term sustainability, it is important to instigate strategies and actions to promote the importance of data and its management, to form appropriate coordination mechanisms at national, regional and global levels, and maintain training and capacity development.

A number of complementary activities are therefore proposed. These are briefly described here, and elaborated on in Appendix D.

In –country level

1. **Raise the in-country profile of NMHS data management**. Many, if not most, NMHSs do not regard data management as a core and essential function that needs adequate investment. Therefore, without engagement within the NMHS (and its country's Government), commitment to *ongoing* CDMS and data rescue activities with NMHS resources will likely not be achieved. Part of this engagement is to demonstrate that useful and valuable climate services which support DRR, agriculture, and revenue generation depend on easy access to well-maintained data.

2. **Ongoing training and competency development**. A core function of the implementation teams visiting countries to install CDMS and instigate data rescue programs is to ensure that a culture of training and broad-based skill development is established and maintained. Experience gained from former projects, along with WMO's capacity-building program, are expected to form the basis for establishing such a culture.

3. **Work to establish partnerships with NMHSs in neighbouring countries**. Especially where CDMS and data rescue projects are rolled out across a number of countries simultaneously, attempts should be made to encourage information sharing with neighbouring countries and a community of practice approach.

Regional level

4. **Form partnerships with Regional Centres of excellence.** The aim here is to encourage regional Universities and other educational institutions to develop training courses in climate data management, and to explore the possibility that suitably-qualified staff can be embedded within regional bodies such as ACMAD in Africa and SPREP in the Pacific to assist NMHSs.

5. **Regional coordination activities.** Encourage the development of a specific function regarding support to national CDMS implementation within Regional Climate Centres. Also, collaborate with and work to build on the existing linkages and knowledge available through regionally-based data rescue collaborative initiatives such as INDARE and WACA-DARE, through workshops and seminars

6. **A pool of relevant experts able to provide support as needed to CDMS and data rescue** (thereby removing the issue of single-person dependencies – a problem even in developed countries). WMO already harnesses the skills of suitably qualified personnel for specific CDMS and data rescue projects. In addition, at regional and national level, capability could be developed from the regional workshops and seminars referred to in the Implementation Plan above, and/or from Training courses identified under (4) above.

Global level

7. **Encourage development of technologies and processes for crowd-sourced digitisation and imaging.** By harnessing the people-power offered by having large numbers of enthusiastic amateurs perform digitisation, initiatives such as OldWeather have demonstrated that large amounts of data can be transcribed from paper records into databases.

8. **Open Source CDMS development, leading to a convergence of technologies over time.** A proposal has been developed for an Open Source CDMS that is compliant with WMO No 1131, with development hosted by the OpenWIS Association. In this Plan, funding for development of a prototype OpenCDMS is recommended (subsequent development into a fully functional system would be funded separately from this Plan). Developing OpenCDMS involves a considerable initial outlay, but would be expected to lead to significant future cost efficiencies, especially if the scope is expanded to include data from other environmental domains such as Hydrology. The proposal would be to support existing CDMS in D & LDCs via a funding stream and the other elements

outlined here, while gradually transitioning into supporting implementation of OpenCDMS.

9. **Ongoing collaboration with global initiatives.** The scope for collaboration with well-established initiatives already involved in data rescue such as ACRE and IEDRO should be explored. If long-term collaborative arrangements can be established, the prospect for more efficient data rescue based on economies of scale may be achievable. Similarly, ongoing collaboration with the development teams of the various "franchise" CDMS developers will be necessary to achieve an orderly transition into more generic Environmental DMSs.

10. **Keep abreast of, and build upon, the outcomes of broader WMO cross-programme data- and information management modernisation approaches and infrastructure**.

    A number of WMO initiatives are currently under way, aiming to improve the efficiency of cross-domain data- and information management. These include, for instance, adopting Cloud-based approaches to hosting both data and technologies. Similarly, collaboration should be sought with work underway under the European Union-initiated Copernicus programme. There is also a need to harness the benefits of emerging technologies such as potentially-improved optical character recognition software.

Appendix A – Current state of global CDMS and Data rescue activities

I – CDMS

Recognising the importance of secure, well-managed and accessible data, WMO implemented its CLICOM project during the 1980s to provide capabilities to Member countries for, among other things, storing, digitizing, quality controlling and analysing climate data, as well as performing a number of other climate data management functions. While highly successful at the time, technological change gradually led to CLICOM becoming obsolete. In the late 1990s, WMO launched a new Climate Data Management System (CDMS) initiative "to meet the growing needs for improved data management capabilities and data services". In response, a number of new CDMSs were developed by Members, and installed in various countries supported by WMO VCP, bilateral funding initiatives, or country-funded project initiatives.

A number of these CDMS were evaluated by WMO and installed in multiple countries. The most popular of these so-called "franchise" CDMSs in terms of the number of installations are CLIDATA, CLIMSOFT, CLISYS, CliDE and the Hydrology-oriented system MCH, while a considerable number of countries still deployed the outdated CLICOM system. Other "franchise" systems included CLIWARE and CLDB. Some 40 respondents still stored their climate data on spreadsheets or flat files.

The above information was collected via a survey[3] of CDMS capability carried out by ET-CDMS and published in 2013. The survey further revealed that while nearly all responding countries (response rate 137 out of 190 countries) had computer-based data management systems, approximately 40 percent of Members indicated that their CDMS was either not fully operational or not operational at all. The main reasons given were: software problems, lack of training, hardware problems and infrastructure issues. Other issues identified were: lack of budget, lack of computer science and climatological knowledge. Overall, many countries indicated that their capability to effectively manage their climate data was unsatisfactory.

Since 2012 most of the "franchise" CDMSs have been upgraded, and in some cases (particularly CliDE), extended to other countries. A few countries have used their CDMS to digitize parts of their hard copy records. However major gaps in data management capacity remain in many countries, and there have been concerns expressed about the long-term sustainability of their CDMS.

In 2014 the ET-CDMS published its specifications for CDMS as WMO No 1131. This document, which was endorsed by Cg XVI, specifies the functionality for a CDMS (either as part of the CDMS or readily accessible to it), categorized according to whether the functionality was required (i.e. mandatory for proper performance of climate data management), recommended, or optional. It is believed that no current CDMS in a Member country, franchise or otherwise, fully meets even the required functionality, although a survey is being developed by ET-CDMS to assess this further. Despite this, existing "franchise" CDMS are capable of supporting sufficient of the mandatory needs (data entry and ingest, secure archival, access, basic quality control, etc) as to represent

[3] The survey was carried out over the period 2010-12, and published in 2013, so details may well have changed. However it is believed that the general conclusions drawn from the survey remain current.

a viable option for data management for those countries currently without an adequate data management capability.

A problem with the existing breed of CDMS is that they generally differ from CDMS to CDMS in their technologies, data models and functionality. This poses a challenge to general attempts to implement and support CDMS by WMO in a sustainable way. Future convergence of technologies would therefore be an advantage.

Noting the need for such convergence, and the need to add the extra functionality to meet the requirements of WMO No 1131, an OpenCDMS project has been proposed, to be carried out in collaboration with OpenWIS. This project is seen as a multi-year endeavor that builds upon existing "best of breed" functionality, develops new functionality as needed, emphasizes Open Source technologies, and evolves with time as a truly open-source initiative. Indicative costs indicate an initial investment of around 25M Euro over five years would be required to develop and test a fully-compliant OpenCDMS[4] module. Despite the daunting size of such an investment, when averaging this out over the considerable number of countries that would benefit, the investment per country becomes one of several hundred thousand Euro per country as a once-off development cost. Such a development would greatly enhance the sustainability and efficiency of NMHS data management in the longer-term, and accordingly an allowance to construct a prototype system has been factored into the current Resourcing Plan. Finally, recognizing the efficiencies of a cross-domain approach to data management, it is proposed that the future evolution of NMHS-based CDMS be towards management of a wider domain of environmental variables, especially hydrology and marine data, i.e, that a CDMS be understood as an Environmental DMS (EDMS). In principle there seems no obvious barrier to adopting such an approach. Beyond that, harmonization of the future development of CDMS or EDMS with broader attempts within WMO (such as the CBS Task-Team Information Management) to modernize data management infrastructure at national, regional and global scales is required. This should also include collaboration with initiatives such as the Copernicus-initiated projects aiming to provide greater access to climatically-significant global datasets

II - Data Rescue

The need to extend historical data-sets backward in time to support climate research and related activities such as historical reanalysis, along with the recognition that the hard-copy data holdings in many countries (especially D & LDCs) are at high risk of damage or loss, has been a strong driver for a number of data rescue initiatives, at national level, regional level and globally.

[4] A Copernicus initiative to develop a centralised data store for all Climate ECVs is not regarded as an alternative to OpenCDMS, because the latter represents a global archive of ECVs to support global climate functions, whereas NMHSs have a need to store and manage a wider variety of data for their in-country service provision.

Part I Global

The **Atmospheric Circulation Reconstructions over the Earth (ACRE)** initiative is an ongoing global undertaking seeking to find, recover, quality control and consolidate global historical surface and some marine weather observations in support of developing reconstructions of the earth's atmospheric circulation over the past 200-250 years. ACRE is led by a consortium of nine core partners from various parts of the world, all with an interest in historical climate data and research. A core part of their activities is data rescue, and they have abundant expertise and influence in this area. They have established, or participated in the establishment of, a number of regional initiatives such as INDARE (refer below). ACRE are represented on the WMO Commission for Climatology's ET-DARE by Dr Rob Allan (Hadley Centre, UKMO)

The **International Environmental Data Rescue Organisation (IEDRO)** is a globally-focussed initiative specifically dedicated to historical (climate) data rescue. Their stated mission is to "locate, rescue, image, digitize, and share historic climate data, enabling developing countries to adapt and mitigate the effects of climate change". As part of this, they initiate Data rescue activities in developing countries, including assisting with the purchase of equipment such as scanners, training of data rescue staff, and carry out digitisation of pre-scanned images. IEDRO are represented on CCl's ET-DARE by Dr Rick Crouthamel (Executive Director, IEDRO).

CCl Expert Team-Data Rescue (ET-DARE). The ET-DARE represent WMO's attempts to coordinate and inform data rescue efforts among WMO Member countries, and in recent years this team has been responsible for the development of the I-DARE data portal (see below) and an updated set of Guidelines on Data Rescue. The Team has also led, or participated in, a number of regional and national data rescue initiatives, such as WACA-DARE (see below). Links with the activities of major players in Data rescue ACRE and IEDRO are represented via the contacts identified above.

I-DARE. Recognising that there were numerous data rescue activities in train on global, regional and national scales, a pressing need was identified to better coordinate these activities. For this reason the ET-DARE oversighted the development of an international data rescue portal, I-DARE (see https://www.idare-portal.org/) , whose aim is to "provide a single point of entry for information on the status of past and present to- be-rescued data worldwide, and (on) data rescue projects, on best methods and technologies involved in data rescue, and on metadata for data that need to be rescued". Development of the Portal, and its ongoing maintenance, is funded by WMO, and I-DARE is hosted by the Netherlands Meteorological Service (KNMI).

For I-DARE to fully represent data rescue needs and priorities in a timely fashion, it is important that Member countries provide information on their to-be-digitised data needs, and details of any data rescue activities being undertaken. For this reason, WMO has requested Member countries to nominate national I-DARE contact personnel, representing the point of contact on data rescue matters for that country.

Other. There are data rescue components in various other international initiatives, including the International Surface Temperature Initiative (UKMO), Global Precipitation Climatology Centre (DWD) and NCEI (USA). These initiatives do not in general carry out data rescue themselves.

Other domains.

The focus of I-DARE is presently on atmospheric data, but as mentioned in the body of the main paper, it could relatively easily be scaled to accommodate data from different domains. *Marine data*. The International Comprehensive Ocean-Atmosphere Dataset (ICOADS) managed by NOAA is an extensive archive of surface marine climate data. The sources for the historical data are very extensive (refer to Woodruff *et al* 2014 for a detailed summary), and have included a great deal of data rescue, including imaging and digitization work, of sources such as ship's logs, diaries etc. Much of this data rescue work was funded within the USA under a public-private Climate Data Modernisation Program (for which funding ceased in 2012).

Hydrological data. There do not appear at present to be any significant data rescue projects within the Hydrology domain, however there is a recognized need to undertake rescue of data for various hydrological variables, such as river discharge.

Part II Regional

Regional collaboration initiatives. WMO has, over the years, led or participated in a number of initiatives to enhance regional collaboration on Data rescue efforts. These efforts have frequently been done in the context of extending data series to assist in the development of a series of Climate Change Detection Indices developed by CCl. Whilst not specifically carrying out or funding data rescue projects, they have provided a mechanism for needs-gathering and collaboration between blocks of countries participating in the regional initiatives. Examples of such regional initiatives include MEDARE (covering Mediterranean countries), SACA-DARE (Southeast Asian countries), WACA-DARE (West African countries) and INDARE (Indian Ocean countries, including small-island developing states).

Copernicus. In 2016 the European Union announced through its Copernicus programme a series of projects aimed at enhancing data management-related activities in support of its Climate Change Service. One of the focus areas was data rescue, and a four-year project (to commence in 2017) has been tendered to develop a high-end international data rescue portal and registry IT infrastructure with new data rescue tools (including some crowd-sourcing capability). There is obviously a need for close collaboration between WMO's initiatives in data rescue (especially I-DARE) and those of Copernicus, although given the four year time-frame of the latter, and the focus of the latter on developing infrastructure rather than carrying out data rescue, any merging of capabilities is most likely to take place outside the five-year time-frame that is the focus of this Plan.

Part III – National and bilateral initiatives

A number of bilateral and multilateral data rescue initiatives have been carried out or commenced in D & LDC utilizing funding provided by donor countries through Aid-funding etc. An example is the data rescue initiatives, with some digitisation, carried out in Pacific countries using funding provided by the Australian Government between 2009-13. This built on earlier data rescue work (2005-06) in the Pacific, which aimed at securing hard copy records and producing inventories, carried out via a bilateral initiative by the Australian Bureau of Meteorology and New Zealand's NIWA.

Part IV - Other -Crowd-sourcing & Intelligent Character Recognition Techniques

The emphasis in many data rescue projects is on securing, imaging and digitizing monthly, daily and sub-daily surface observations. For large countries with many historical observations, the time and cost to do even this much may be prohibitive. Moreover, such efforts typically do not include the imaging and digitisation of autographic charts or of contextual metadata. It could be argued that, in the longer term, crowd-sourcing approaches are needed to make significant inroads into the large repositories held in some countries.

There have been a number of demonstrably-successful crowd-sourcing digitization initiatives . Examples include OldWeather (which focused on digitizing ship's logs from early in the 20th Century), and Australian-based local citizen-science initiatives at the University of Queensland (Weather Detective) and at the Bureau of Meteorology's South Australian Office. However there currently appears to be no standard technology or guidelines for establishing crowd-sourcing projects for digitization or imaging.

Various attempts have been made to utilize Intelligent Character Recognition software as a means of automatically transcribing data. To date, these have not proven very successful at consistently interpreting hand-written logs and field books. However there are hopes within the Data rescue community that emerging machine-learning-based techniques may lead to significant future improvements in this area.

Appendix B – Lessons learned

As described in Appendix A there have been a number of successful CDMS implementation and data rescue projects in Developing and Least Developed countries (D & LDCs) on a regional scale. These have been instigated through either regional initiatives such as WACA-DARE in West Africa, or aid-funded projects such as the CliDE CDMS and data rescue activities in the Pacific funded by the Australian Government and more recently by the GFCS. These experiences have enabled a body of knowledge to be assembled about the particular issues applying to such work in D & LDCs and how to address them. An awareness of these issues is crucial to the successful implementation of CDMS-Data rescue projects.

While not attempting to provide comprehensive guidance, the main issues and learnings from this experience may be summarized as follows:

1. **Profile and promotion**. While the long-term sustainability of CDMS and Data rescue in D & LDCs in particular may be assisted by international efforts and funding, this will be insufficient unless there is support from within the NMHS and the parent government. As described in Appendix D this requires a clear demonstration of the benefits of supporting data rescue and ongoing efficient data management. Such benefits might include, e.g., how accessible climate data can help efficiently answer a minister's question about the frequency of heavy rainfall or drought, and also the potential to raise revenue by providing value-added advice or products based on the climate data.

2. **Staffing**. Many NMHSs in D & LDCs have few staff, with multiple responsibilities. Therefore their ability to perform ICT maintenance work and activities such as digitisation will be limited. Moreover, trained technical staff are generally in short supply, and should they leave the NMHS, often leave a sizeable gap. This is why when training in Data rescue and CDMS is conducted, it needs to be done in-country so that several staff can be trained at once, in the NMHSs' working environment. An additional benefit of this approach is that, should an NMHS in a neighbouring country need help, availability of a pool of staff within an NMHS means that someone can provide that assistance without significant disruption to the providing NMHS' own activities.

3. **Intra-office collaboration**. Weather services such as forecasting and warning provision make use of real-time or near-real-time data from a variety of sources. By contrast, many climate services are based on time-series data on a non-real-time basis. It is desirable to ensure, however, that the data-sources utilised for weather services, such as SYNOP messages and data from AWS, also flow into the climate archive. Therefore it is strongly recommended that within the NMHS, planning discussions take place to ensure that either the incoming near-real-time data will flow into the CDMS, or that such data can be readily accessible to the CDMS via other means.

4. **Communications infrastructure**. Communication infrastructure in D & LDCs is frequently not robust, and there may be insufficient internet bandwidth to support some of the emerging ICT capabilities such as accessing technologies from a Cloud

environment. In some Pacific countries bandwidth is insufficient even to back up their CDMS efficiently. Unfortunately there appears to be no easy or cheap solution to this problem at present.

5. **Data Policy**. Although Resolution 60 mandates that member countries allow the free exchange of their historical climate data, in practice in many D & LDCs this is not currently permissible, whether because of the NMHS' own policies or overarching government policy. This has not only caused problems in accessing data for regional or global climate-related projects and research, but from a data management perspective rules out, or makes it harder to implement, particular solutions such as backing up data in other countries, or centralised regional digitization-imaging projects.

6. **CDMS design**. Given the skills shortages in many D & LDCs, it is imperative that the design of the CDMS is as simple as possible, in particular with intuitive, easy to use interfaces, clear and simple user manuals, and robust checks and balances to ensure data cannot be lost, corrupted or inappropriately accessed or altered.

7. **CDMS design (2**). In the absence of a purpose-built CDMS, many countries store their data on Excel spreadsheets or similar, which are functional but less efficient than CDMSs, and are prone to such issues as high exposure to computer viruses, backward compatibility issues, electronic storage limitations, and other problems.

8. **CDMS design(3).** There are a number of "franchise" CDMSs currently deployed globally (where in this sense "franchise" is understood to be a particular CDMS installed in several different countries). Examples of such "franchise" CDMSs include CliWARE, CliSys, CliData, ClimSoft, CliDE and MCH. Some of these systems are built around proprietary technologies such as Oracle databases, some are Open Source. The proprietary-based systems require the payment of regular licensing and administration fees, which constitutes an additional overhead to data management support. Where the NMHS concerned is prepared to pay these charges, this need not be an issue; however it is a decision for WMO and partners as to whether it is prepared to subsidise these charges on an ongoing basis.

9. **CDMS implementation and Training**. The initial visit to an NMHS to implement a CDMS should be of sufficient length to install the CDMS, configure it within the existing NMHS IT infrastructure and in accordance with the NMHS workflow processes, migrate data from existing systems, and conduct training (noting the desirability of training multiple staff). In general the time required will depend on the NMHS needs and staff availability. The costs provided in Appendix C allow for a two week initial visit.

10. **Data Rescue**. Most NMHSs do not have the infrastructure to adequately secure their hard copy records, which require storage in well-labelled acid-free boxes in a dedicated storeroom on shelves (refer WMO No 1182, Appendix 3). This is why the costings in Appendix C allow for such equipment. Some NMHSs do not even have the physical space for such storage, although some NMHSs have solved this problem by, e.g., acquiring a shipping container, or storing in a public building outside the NMHS such as within a university or the country's national archives.

11. **CDMS-Data rescue project synergy**. Except when and where data rescue activities have already been commenced under a regional initiative, it is recommended that a

climate data specialist accompany the CDMS installation technician to initiate data rescue activities, including assisting staff to develop inventorying, storage and imaging and digitisation activities. Among other synergies, this should allow for a program of digitization into the CDMS to be established at the outset. Guidance and recommended practices for establishing data rescue projects are described in detail in the updated version of "*Guidelines for best practice on data rescue*" (WMO No 1182), and on the I-DARE portal at https://www.idare-portal.org/

12. **Follow-up visits** are essential, to ensure that staff are using the software appropriately and that any issues or problems can be addressed. Ideally this should take place in the same year of installation, although experience suggests that in practice many NMHSs will not need this. In any case, however, follow-up visits should take place within two years of installation. This will provide an opportunity to conduct refresher training and assistance for staff in both CDMS operation and data rescue activities, and implement and explain any enhancements and modifications to the CDMS. The costings in Appendix C allow for such a follow-up visit.

13. **Timely support from the project team is essential**. In any case, the CDMS installation team must at all costs be ready to provide support to countries experiencing difficulties with their CDMS, as experience has shown that if they cannot readily resolve the problems, they will likely cease using the CDMS and revert to spreadsheets or older but more familiar technologies such as CLICOM. In that case, a large part of the investment in the CDMS is lost. For this reason, the costings in D allow for a number of contingency visits annually, but equally, there is need to ensure that the CDMS design and implementation team are able to respond in a timely fashion.

14. **Timing of in-country travel**. In planning such implementation and follow-up visits it is important to allow for seasonal weather and climate limitations. Scheduling a visit to a tropical small island developing state during the cyclone season, for instance, runs a high risk that NMHS personnel will become unavailable due to the need to prioritise operational forecast-warning duties.

15. **Establish communities of practice**. In regional initiatives, it is desirable to establish a Community of Practice approach whereby online communication is enabled between personnel in countries participating in the initiative, and between such staff and those in CDMS-Data rescue project teams. The benefits of such an online community is to provide a forum for solving problems that individual countries might face, and to advise of recent developments, upgrades etc.

Appendix C – Costings and assumptions

Below are first-order cost estimates required to initiate CDMS implementation and data rescue activities, and maintain them for five years in D & LDC countries identified as high priority. The approach taken here is to form an estimate of the average cost per country based on past experience, and assuming that, to a reasonable extent, such an estimate would be extendable to other countries and regions. Details of costs in individual years, and for each activity, are provided in **Attachment 1 – costing breakdown.**

The costings here are based on actual CDMS deployment and data rescue activities in Region V, with some additions based on experience in Region I. We therefore believe these provide a sound basis for estimating costs in other D & LDCs.

The per-country cost estimates here apply to a full CDMS-Data rescue implementation cycle, but in many cases actual costs would be lower. In many cases some of the activities are already underway (e.g. an initial assessment has already been conducted, a CDMS has been installed, imaging equipment provided etc.), and therefore in these countries one would expect the overall cost to be lower. In addition, the costings provided here included a 20 per cent overhead, reflecting the fact that some NMHS may impose such overheads to cover general costs such as providing office space, equipment, etc for project staff.

In the section "5 year breakdown" (pp22-25) year by year costs are estimated, covering staff salaries, travel and equipment, data rescue activities (for both a basic and a more comprehensive program, as described under "Data rescue" on p 26), and the costs associated with staging annual regional workshops.

In the final sheet of Attachment 1 ("Summary Overall", p27), the costs covered in lines 6-15 are aggregates over five years for 15 countries, reflecting Australian experience in the Pacific. From this the average cost per country is calculated (line 18), and this is then applied to a hypothetical program wherein a full cycle of CDMS deployment, training and data rescue is carried out in 35 countries[5] (line 21). This is intended to give a "ball-park" estimate of costs for a dedicated program over five years, and works out at a total of CHF 12,223K or CHF 349K per NMHS

In addition to the in-country activities, estimated costings are also provided (Parts V and VI of "5 year breakdown", and lines 23-24 of "Summary Overall") for maintaining the I-DARE portal as a means of coordinating data rescue activities globally, and extending it to marine and hydrological data; and to develop a prototype of OpenCDMS software, seen as important to the long-term cost-effective maintenance of CDMS.

The overall cost for the in-country program (35 countries) plus I-DARE and OpenCDMS prototype is estimated at CHF 16,473K over the five years. The following assumptions are made:

1. Costings are estimated over a five year period, this being the recommended time required between upgrades of the PCs on which the CDMS is deployed.

[5] The number 35 is an initial estimate based on recent Member requests for assistance in data rescue and CDMS implementation

2. Costs will vary over time. It is assumed that in-country visits to each country would be required in Year 1 (initial implementation, configuration to working environment, and training), and Year 3 (refresher training, upgrades etc.). More general training, information on upgrades and coordination activities would be conducted at once-yearly regional training workshops.

3. In these years it is assumed that 2 people travel to each country, one a data specialist covering data rescue and data management training, and initiating inventory construction and digitisation programs, and one an IT specialist (CDMS installation, configuration, embedding in local infrastructure, training). A two week initial visit (Yr 1) is allowed for to carry out these activities. The IT staff requirement in these years is set at 1.5 full-time staff per Region, to allow for ongoing development work on top of the in-country visits. Experience has shown that this level of staffing could cover activities in up to 15 countries.

4. An allowance for contingencies of three trips per Region is made per year to cover problem-fixing etc, with these being the only travel costs in Years 2,4,5.

5. It is assumed activities in each of three illustrative Regions where many of the identified vulnerable countries are located, are coordinated by a regional project manager representing 0.2 of a full-time person (FTE). That person's role is to ensure project activities within the Region stay on track. Experience in the Pacific indicated that an allocation of 0.2 in years 1 and 3, and 0.1 in remaining years, was sufficient to cover 15 partner NMHSs[6] This averages out at AUD 40K a year (approximately CHF 30K a year)

6. Overall coordination-management is from WMO (0.33 FTE) = 50K CHF (AUD 65K)
7. Administrative support for arranging travel, procurement etc. Allow for funding for up to 0.25 FTE for each Region. This figure would be less if WMO shared some of the load.

8. Salary costs will vary. Salary figures quoted for RA V are likely considerably higher than in other Regions, because project staff were drawn from developed countries, and a project overhead of 20%was assumed. Ideally over time, locally trained champions may take over some of the technical roles, and Regional Climate Centres (RCCs) and other regional centres of excellence may play a role in this regard in future years.

9. In many countries costs would be lower because of pre-existing work. For instance in Region V, because CliDE is already in place, actual costs would be less than quoted in Year 1 However a data expert would still be required to perform inventory work and establish digitisation and imaging projects in most countries.

10. Travel: In Region V, an indicative cost is AUD 5K AUD (= 4K CHF) per person for one week's travel and accommodation. An equivalent figure in RA I would be approx. 3K for someone travelling from Europe.

[6] An independent estimate assuming 0.2 FTE per year from WMO@ 150K p.a. gives a very similar value of CHF 30K.

11. Data Rescue: In the Spreadsheet "*Data rescue*"", two scenarios for data rescue are presented:
    - Basic Data rescue set-up with some digitization but no imaging[7], covering the following per country, and based on experience in the Pacific and Africa:

Boxes, freight	5K
Shelving	11K
Inventories	2K
Digitisation	15K
Sorting	3K
Cleaning of records	1K
Overhead	7K
TOTAL	**43K**

    - A more advanced Data rescue program, covering an assessment visit, increased digitization (30K over three years), imaging (35K over three years), and 12K for imaging equipment that includes a dedicated PC for storing the images. This more advanced program works out at 126K per NMHS.
12. Data Rescue: The digitization costs could vary considerably between countries, depending on:
    - Whether all daily and sub-daily data are digitized;
    - Whether externals are hired to perform digitisation or NMHS staff (figures quoted assume external contractors are hired).
    - The number of stations with data to be digitized;
    - The length of record;

    As examples within RAV:

    - a current digitization project in PNG aiming to digitise pre-independence (1973) synoptic and rainfall data using casual staff, including the provision of equipment and IT support, was costed at approx. AUD 15K;
    - a 2009 project to digitize all Vanuatu's daily and monthly synoptic and rainfall data, plus six-hourly temperature and dewpoint observations (seven stations), cost approx. AUD 16K, or about AUD 18.5K in current values.
        - However in some of the larger countries the resources to fully digitize all the historical records may be many times larger. As an indication, in 2005 it was calculated that 790 person-years would be required to rescue all data in the DWD archives.

13. Regional training and coordination workshops. Required for ongoing training, updates, and to maintain enthusiasm and collaboration. Assume 1 each per Region per year to cover Africa, the Indian Ocean countries and the Pacific, at a cost of approximately 100K each. This training would contribute to the development of a core of specialists, able to provide support and maintenance as

[7] Reflecting the way this was done in the Pacific. In Africa, greater weight is being given to imaging.

required for other Data rescue projects and CDMS maintenance. Cost AUD 500K per Region.

14. I-DARE portal: This provides an ongoing means of monitoring requirements for, and progress made, in recovering and digitizing climate records, and identifying gaps and priorities. I-DARE also hosts documentation on best-practice in data rescue, and related tools and technologies. To maintain I-DARE at present costs approximately 5K Euro (5.2CHF) p.a. To add functionality to record digitized data and extend to marine and hydrological data is estimated to cost an additional 10K Euro, and increase ongoing maintenance costs to approximately 7.5K Euro p.a. This works out at a total cost of AUD **65K** (approx. CHF 50K) over 5 years to establish and maintain an I-DARE with expanded functionality.

15. Historical data from some NMHSs (especially those with a colonial past) is often held in the archives of other countries, but the existence of such data, and exactly what data is held, is often not known. Therefore a useful step towards populating I-DARE would be for WMO to encourage Members to provide details of any data they hold from other countries

16. OpenCDMS. A longer-term vision for cost-effective maintenance of CDMS is to converge technologies so that eventually just one software system needs to be maintained, not several as at present. The costings given on the worksheet "5 year breakdown" are initial estimates of what it would cost to develop a prototype capable of carrying out at least the basic data management functions according to WMO No 1131.

Attachment 1 – Costing breakdown

Year by year breakdown

Unless otherwise stated, all values in this and the following worksheet are in thousands of dollars AUD

Part I - Staff costs

Year	Period	No staff	Salary cost	Over-head	Total Cost	
1	Annual	1.5	281.25	56.25	337.5	CDMS installation
1	Annual	1.0	187.5	37.5	225	Data rescue project commencement
1	Annual	0.2	47.2	9.4	57	regional coordination
1	Annual	0.25	40.3	8.1	48	Admin support
2	Annual	1.0	187.5	37.5	225	CDMS support and development
2	Annual	1.0	187.5	37.5	225	Data rescue support
2	Annual	0.1	23.6	4.7	28	
2	Annual	0.2	32.2	6.4	39	
3	Annual	1.5	281.25	56.25	338	Follow-up visit
3	Annual	1.0	187.5	37.5	225	Follow-up visit
3	Annual	0.2	47.2	9.4	57	
3	Annual	0.25	40.3	8.1	48	
4	Annual	1.0	187.5	37.5	225	
4	Annual	1.0	187.5	37.5	225	
4	Annual	0.1	23.6	4.7	28	
4	Annual	0.2	32.2	6.4	39	
5	Annual	1.0	187.5	37.5	225	
5	Annual	1.0	187.5	37.5	225	
5	Annual	0.1	23.6	4.7	28	
5	Annual	0.2	32.2	6.4	39	
TOTALS			**2,405**	**481**	**2,886**	

Assumptions

1. In most countries, an assessment visit would also be required (not needed in RA V). It is assumed this would be undertaken by WMO staff or experts, so staff time is not costed here
2. 1.5 Full-Time Equivalent (FTE) IT staff per region in years 1 and 3, for in-country CDMS installation, maintenance, and follow-up visit in Year 3, while still allowing for some development work.

3. 1 FTE data person to support data rescue inventories, training in data management etc.
4. Assume 0.2 FTE for project oversight in Years 1 and 3, 0.1 in other years.
5. An allocation of 0.25 FTE is assumed to cover Administrative costs associated with travel, procurement etc., but 0.2 in Years 2,4, and 5 reflecting reduced travel, and less again if WMO can assist.
6. Overall coordination from WMO to be provided at 0.33 FTE at P4 level (CHF 50K p.a.)

Year	**Part II - Travel / equipment REGION V**	Assumption: AUD 5K per person for a one-week trip
1	300	Assumes 2 people x no countries (15) x 2 week
1	15	Assumes total 3 contingency trips per year
1	63	Overhead
2	0	
2	15	Assumes total 3 contingency trips per year
2	3	Overhead
3	150	Assumes 2 people x no countries
3	15	Assumes total 3 contingency trips per year
3	33	Overhead
4	0	
4	15	Assumes total 3 contingency trips per year
4	3	Overhead
5	0	
5	15	Assumes total 3 contingency trips per year
5	100	CDMS hardware PC replacement
5	23	Overhead
TOTAL	**750**	

Assumptions

1. Assume 5K per person-trip per week
2. Year 1 is implementation, Year 3 is follow-up visit for upgrades, refresher training
3. An assessment visit is included in the Data rescue estimates (refer next sheet), and not included here.
4. All years have contingency of 3 trips per year for urgent maintenance and other contingencies
5. Equipment includes PC replacement in Year 5

Part III - In country data rescue activities (Per country total over 5 years)

Refer next sheet

	Boxes, freight	Shelving	Inventory	Digi-tisation	Sorting	Cleaning	Assessmt. visit	Imaging equip	Imaging labour	Dedicated DR PC	Over-head	TOTAL
Basic DR activities	5	10	2	15	3	1.5					7.3	**43.8**
Recommended DR activities	5	10	2	30	3	1.5	7.5	8	35	4	20	**126**

Approx. 600K AUD initially estimated for basic data rescue activities across the 15 countries (40K per country)

Adding assessment visit, extra digitization and introducing imaging etc. (including a dedicated PC for the images) increases to **1,890K** (126K per country) - refer next sheet

Part IV - Regional training/Coordination meetings

One per year x 5 years

@ 100K Total **500K** Covers updates and refresher training CDMS and Data rescue, Community of practice activities etc.

Part V - I-DARE support (over 5 years)

	Initial	Ongoing p.a.	Ongoing p.a. with all three
Maintenance		6.85	
Upgrade	6.85		
Extension	6.85		
Total			10.28

Note: 5K Euro ~ AUD 6.85K

So current maintenance of I-DARE is approx. 6.85K AUD per annum,
plus it is estimated that upgrade and extension would cost around 13.7K
and then around 10.3K AUD per annum to maintain the updated site.
So over 5 years total cost would be (5*10.3+ 13.7) = **AUD 65.2K ~ CHF 50.1K**

Part VI - OpenCDMS prototype

Costings below cover development of an Open Source Climate Data Base Management System ready to install, well documented with a manual and training materials

YEAR	Cost (AUD)	Cost (CHF)	Activities
1	1,700	1,300	
2	1,900	1,400	
3	1,900	1,500	
TOTALS	**5,500**	**4,200**	

DATA RESCUE

Estimates per NMHS, in AUD

	YR 1	YR 2	YR 3	
Initial assessment	7.5	-	-	Not all countries need this. This figure applies to ones that do.
Shelving, refurb	10	-	-	
Sorting	3	-	-	
Boxes and freight	5	-	-	
Inventorying	2	-	-	
Digitization	10	10	10	
Imaging - Equipment	8	-	-	Scanners, camera, camera stands, lights, etc, PC
Imaging- labor	15	10	10	Based on 2 people per year hired, 3 in Yr 1
Dedicated PC for image storage	4	-	-	
Record cleaning - equipment	0.5	-	-	Experience suggests not all countries need records cleaned, say one in 5. Costs spread on that basis
Record cleaning - labor	1	-	-	i.e. Equipment 2K, labor 5K
Overhead	12	4	4	
TOTALS	**78**	**24**	**24**	**126**

(or **1,890K** over 15 countries)

SUMMARY OVERALL

All values in multiples of 1,000

(The 2nd column in each case if for the full data rescue programme, as per worksheet "Data rescue")

	AUD		**EURO**		**CHF**	
	5 yr totals		5 yr totals		5 yr totals	
Staff	2,886	2,886	2,107	2,107	2,220	2,220
Travel	750	750	547	547	577	577
Data rescue	600	1890	438	1380	462	1454
Regional training/coordination	500	500	365	365	385	385
Hardware repl	100	100	73	73	77	77
TOTALS (for 15 countries**)**	**4,836**	**6,126**	**3,530**	**4,472**	**3,720**	**4,712**
Add overall coordination	65	65	47	47	50	50
Contingency 10 per cent	490.1	619	358	452	377	476
Grand Totals (15 countries**)**	**5,391**	**6,810**	**3,888**	**4,923**	4,147	5,239
Per country	**359**	**454**	**259**	**328**	**276**	**349**
(aggregated over 5 yrs)						
Extended to 35 countries	**12,579**	**15,890**	**9,071**	**11,488**	**9,676**	**12,223**
I-DARE (65K/15 countries)	65	65	47	47	50	50
OpenCDMS	5,500	5,500	4,014	4,014	4,200	4,200
TOTALS for all activities (35 countries + I-DARE+ OpenCDMS)	**18,144**	**21,455**	**13,132**	**15,549**	**13,926**	**16,473**
			Exch rate 1 Euro ~ 1.37 AUD		Exch rate 1 CHF ~ 1.30 AUD	

Not considered

Autographic charts digitization

Develop crowd-sourcing module

Inflation

Attachment D – Complementary activities to enhance sustainability

In-country

1. **Raise the in-country profile of NMHS data management**. While a significant upfront investment to implement CDMS and data rescue activities is justified in the initial stages, to be sustainable it is important that sufficient resources are available within the NMHS to continue to maintain the CDMS (in partnership with WMO and others), and continue data rescue work, as in many cases the funding required will be well short of what is needed to carry out all data rescue, imaging and digitisation. It is therefore very important to raise the profile of data and its management within the NMHS (and with the country's Government). To this end, it is recommended that efforts be made to demonstrate that useful and valuable climate services that support DRR, agriculture, and revenue generation depend on easy access to well-maintained data. Examples include being able to demonstrate that responses to ministerial questions about drought or flood events are much easier to provide if the data are readily accessible and well-managed, and similarly, that well organised data archival helps generate revenue through the provision of value-added climate services. To assist with this, it is recommended that DRR as part of its strategy (cf. Resolution 10 (Cg-17)) highlight how its activities depend on the availability of historical climate data, and that WMO develop guidance material aiming to help NMHS staff make this case.

2. **Ongoing training and competency development**. A core function of the implementation teams visiting countries to install CDMS and instigate data rescue programs is to ensure that a culture of training and broad-based skill development is established and maintained. Experience gained from former projects, along with WMO's capacity-building program, form the basis for establishing such a culture. For instance, in-country visits enable several staff at a time to be trained in the maintenance and use of CDMS and in data rescue techniques. This not only reduces the risks associated with single-person dependencies, but enables NMHSs to provide support and assistance where needed to NMHS contemporaries in neighbouring countries.

   Such training should also extend to **General capacity building in climate (and general) data management**, such as backup strategies, security, technological migration. This would involve the development of a climate data management training course (and some countries, such as Météo-France have already developed suitable training material), and it is recommended that this be part of WMO's general capacity-building strategy.

3. **Work to establish partnerships with NMHSs in neighbouring countries**. Especially where CDMS and data rescue projects are rolled out across a number of countries simultaneously, attempts should be made to encourage information

sharing, even resource-sharing, between neighbouring countries. For instance, should an NMHS strike a problem with their CDMS, the problem may easily be solved via consultation between data management peers in neighbouring countries. The costings in Appendix C allow for a number of regional workshops that bring data managers with common CDMS together on an annual basis for refresher training, and these could also serve to enhance such collaborative activities, and to foster communities of practice. Experience in the Pacific has shown that certain individual NMHSs may emerge as regional "champions" in that regard.

Regional

4. **Form partnerships with Regional Centres of excellence.** Examples include with universities, regional bodies such as ACMAD, SPREP etc. The aim would be to encourage training and development of the next generation of climate data managers, perhaps by assisting educational institutions develop curricula on climate data management. While the practicality of utilising suitably qualified staff embedded within regional bodies such as ACMAD needs to be assessed, in principle such staff may be able to provide assistance to NMHSs at much lower cost than if such expertise had to be continually sourced from developed countries.

5. **Regional Coordination activities.** Currently Regional Climate Centres do not have a dedicated function regarding support to national CDMS implementation and operation, but they potentially could serve as a base for technical staff to support CDMS and data rescue within a Region, coordinate training activities etc., and it is suggested that investigations be carried out as to what extent this is feasible. Similarly, regionally-based data rescue collaborative initiatives such as INDARE, WACA-DARE, SACA-DARE have already collated a good deal of knowledge on needs in individual countries, and already encourage the sharing of expertise among countries (as per (3) above).

6. **A pool of relevant experts able to provide support as needed to CDMS and data rescue** This would involve the training of suitably-skilled technicians, plus in-kind contributions from CDMS developers and individual countries. At a regional level, the skills could be developed as part of the annual regional workshops described under Appendix C, and from training courses at regional educational institutions.

Global

7. **Encourage development of technologies and processes for crowd-sourced digitisation and imaging.** Successful initiatives such as OldWeather and others have demonstrated that large amounts of data can be transcribed from paper records into databases by amateur enthusiasts at relatively low cost, but there does not appear to be any generic software or process to enable this to be done on a large scale. A collaborative build process or similar might enable this to be arranged. A prerequisite to crowd-sourced digitisation would be to have the documents imaged.

8. **Open Source CDMS development, leading to a convergence of technologies over time.** While a number of so-called "franchise" CDMSs have been developed and deployed to numerous D & LDCs, in the long-term this "model" has a number of problems. These include: differing data models, architectures and functionality (thereby making it harder to support them); technology that potentially incurs burdensome licensing costs and administration charges; dependence on a limited number of developers (with the attendant problems of key person dependencies, succession planning etc.); and the fact that no existing CDMS is fully compliant with the required functionality of WMO No 1131. A proposal has been developed for an Open Source CDMS that is compliant with WMO No 1131, with development hosted by the OpenWIS Association. This Plan funds the development of a prototype OpenCDMS, with a longer term view of developing a fully-functional version funded from outside the Plan . Once developed and deployed, it is expected that the eventual long-term savings and availability of state of the art CDMS functionality will lead to significant cost reductions over time. This would particularly be the case if the scope of the CDMS were to be broadened to incorporate data from other environmental domains, such as hydrological data. The long-term proposal would be to support existing CDMS in developing countries via other elements of the Plan, while gradually transitioning into supporting implementations of the OpenCDMS.

9. **Ongoing collaboration with global initiatives.** As noted in Appendix A, international initiatives with a strong data rescue focus such as ACRE, IEDRO, ICOADS, etc. have established very considerable experience, credibility, capability, and extensive global reach. While WMO has informal linkages with these groups, consideration should be given to establishing longer-term collaboration protocols with them, including consideration of jointly-funded data recue projects.

10. **Keep abreast of, and build upon, the outcomes of broader WMO cross-programme data management modernisation approaches and infrastructure**.
    The CCl/CBS-led IPET-CDMP and newly-established CBS Task Team-Information Management have been established to improve the efficiency of data- and information management processes across all WMO domains and Member operations. The efficiencies gained from this work, and the harnessing of advances in Information Technology such as Cloud-based data and technology hosting, is expected to lead to significant efficiencies in data and information management over time. For instance, by deploying technologies such as CDMS to the Cloud, so that they function in effect as internet-based software systems, they become more easily accessible, maintenance and upgrades are easier and cheaper [8]to achieve, and access to functionality (and data) remains available in the face of natural, civil or military disasters. This will incur additional costs for implementing and maintaining adequate ICT infrastructure in some D & LDCs, but these skills may be easier and cheaper to acquire.

[8] Noting, however, that there will still be the need to ensure the in-country installation can be sustainably supported and maintained.

Similarly, work underway in other data-focussed domains, including under the Copernicus stable of projects, will also help develop modernised approaches to data rescue and the creation of global archives of Essential Climate Variables, with which the broader WMO initiative may wish to collaborate. Advances in technology such as improved capabilities in optical text recognition (which potentially could greatly improve the efficiency of digitisation) should be identified and harnessed where possible.