

WORLD METEOROLOGICAL ORGANIZATION

**MEETING OF THE CCI EXPERT TEAM ON DATA RESCUE
(ET DARE)**

3-6 November 2014
Geneva, Switzerland

FINAL REPORT

1. Opening

The meeting of the CCI Expert Team on Data Rescue (ET DARE) was formally opened at 9.30 a.m. on Monday, 3 November 2014. On behalf of the Secretary-General of the World Meteorological Organization (WMO), Mr Peiliang Shi, Director, WMO Information System (WIS) Branch, addressed the meeting by highlighting the importance of data rescue. He underlined the opportunity and need to use WIS to make rescued data discoverable, and referred to data rescue-relevant aspects of the UN Big Data initiative, crowd-sourcing mechanisms and private partnerships.

2. Organization of the meeting

The meeting was chaired jointly by ET DARE Chair and Co-Chair, Ms Westcott and Ms Flannery, respectively. The agenda for the meeting (see Annex 1) was adopted with no revisions. The meeting agreed on its hours of work and other practical arrangements. The list of participants is presented in Annex 2.

The agenda and outcome of the joint session of CCI's Expert Team on Climate Data Management Systems (ET CDMS) and ET DARE (Wednesday, 5 November, WMO premises) are provided in Annexes 3 and 4.

All meeting presentations can be accessed at:
<http://www.wmo.int/pages/prog/wcp/wcdmp/Meetings.php>

3. Status of Data Rescue

The team members apprised the meeting on the ongoing Data Rescue efforts worldwide. The meeting appreciated the presentations made by the experts, showing progress made in specific regions and countries as well as through international projects and organisations, such as CliDE, IEDRO and ACRE. A summary of these presentations is provided as follows:

William Wright, Co-Chair of CCI OPACE 1 (Climate Data Management) provided an overview presentation, addressing the wider context of the ET DARE activities, reviewing DARE-relevant achievements during the last CCI intersessional period (2010-2014) and highlighting a couple of important issues to be considered, such as resource mobilisation for, and sustainability of data rescue activities, information sharing as well as needs for technology migration.

Peer Hechler provided an overview of ET DARE-relevant outcomes and recommendations of the 16th World Meteorological Congress, the 66th Session of the Executive Council, the 16th Session of the Commission for Climatology and sessions of the Intergovernmental Board on Climate Services; all of which are reflected in the team's current terms of reference. He presented recent and on-going DARE-related activities in which the Secretariat has been involved recently, including the I-DARE evolution as well as co-ordination of, and assistance to, national and regional data rescue initiatives worldwide.

Nancy Westcott provided an overview on the Climate Data Modernization "Forts" project on extending daily data records backwards to the early 1800's. The project completed the scanning

of 4800 station across the continental United States, with 450 stations keyed, and 325 stations quality controlled, with an additional 25 stations quality controlled from Alaska. These data, including an inventory of the scanned images, and the quality controlled data are archived and available from the Midwestern Regional Climate Center (MRCC <http://mrcc.isws.illinois.edu/research/cdmp/cdmp.html>). To date these data have been used for comparison with other data (tree rings, 1876 weather maps, native American records of climate), and information available from diaries, historical accounts, and historical fiction. These data have also been used for research into weather and climate impacts on: mortality and morbidity on Union Army veterans, forest fires and forest management practices, the frequency of frost bite injuries and the prediction of frost damage on woody forests, and the effect of severe cold on the wine industry. Current ongoing Climate Data Rescue projects in North America include several crowdsourcing activities, one led by Dr Kevin Wood of the University of Washington- NOAA Joint Institute for the Study of the Atmosphere and Ocean in conjunction with the National Archives and the Zooniverse Old Weather project, digitizing ship logs in the Arctic, and a second project in Canada led by Vicky Slonosky digitizing historical Canadian diaries and station data tables.

Richard Crouthamel made a presentation on Application of Historical Climate Data in various sectors. Historic climate data is critical for addressing the needs of agriculture, insurance, public health, and improvements in flood forecasting. Until now, these old paper-based data had to be manually keyed into a digital data base before computer models could use these data for the betterment of humanity. With IEDRO's new digitization program WEATHER WIZARDS, many volunteers will digitize these valuable data in a fraction of the time previously required. The newly digitized data will be automatically quality controlled and entered into the Data Base for the use by the world's scientists, researchers, and educators. He concluded that there is a need for funding to enable the construction of hundreds of keying formats that match these old data forms awaiting digitization before these records are gone forever.

Jose Carlos made a presentation on data preservation and rescue at the Meteorological and Hydrological Service (SENAMHI) of Peru. SENAMHI is undertaking continued efforts for the preservation and restoration of hydro-meteorological and climate data in Peru in line with the WMO Data Rescue program. In 2000 SENAMHI started DARE activity as part of a project called "Improvement of the Forecasting Capacity and Assessment of El Niño Phenomenon for the prevention and mitigation of disasters in Peru", involving a massive hydro-meteorological data entry with improving data management. In 2010 DARE efforts continued in the frame of a project "Automation of Hydro-meteorological Graphic Bands of the National Hydro-meteorological Archive". These led to the digitization of 462,056 bands of meteorological recording instruments. It is necessary to mention that this work has been carried out based on software which was developed specifically for this propose that also includes the curve – data converter. Recently, as part of CLIMANDES I project, DARE efforts include data entry from spreadsheets that include data of the 10 major weather stations in the two pilot regions: Cusco and Junín, having more than 30 years of data. Within the second phase of CLIMANDES II, efforts were undertaken for strengthening all Peru's weather stations which have 30 or more years of data. SENAMHI is promoting data rescue and digitization at the National Technical Archive and its 13 technical offices in the country.

Meaghan Flannery reported on Data Rescue activities in RA-V. Region V (RAV) has been active in DARE activities with many Pacific Island Countries using the Climate Data for the Environment (CliDE) Climate Data Management System (CDMS) for either active digitization of their historical climate record or uploading spread sheets containing this information into the CliDE database. Vanuatu completely digitized all known historical climate records using a team of students. The Australian Bureau of Meteorology (BoM) repatriated over a decade of Dili's historical record to Timor-Leste after preserving and digitizing all data contained in these recovered hardcopy station records. The BoM initiated a program this year (2014) to image historical climate records in preparation for crowdsourcing digitization programs. Currently 60,000 records have been imaged using the recently acquired portable Vanguard archival standard document scanner. The BoM continues to discover and inventory all their historical data holdings archived by National Archives Australia (NAA) data repositories located in Australian capital cities. The outcome of this project is enabling the BoM to extend their historical climate data inventory in preparation for imaging and digitization. There have been several lessons learned from RAV DARE activities including the importance of countries using an expert guidance and training on digitization, the essential requirement to perform user acceptance testing (UAT) in the initial stages of the DARE digitization project to ensure that digitized data is fit for purpose, the need to integrate DARE activities into CDMS project plans, consider the requirement for realistic timeframes and exit strategies to be incorporated into CDMS implementation project plans to address the withdrawal of legacy system use (such as spread sheets), and the implementation of sustainable in country management of the CDMS

David Muchemi presented the Kenya Meteorological Services (KMS/KMD) experience in Data Rescue and digitization of climate records. KMD has large collection of climate data starting from the 1900s. Most of this data is on paper and there have been many initiatives to rescue this data as far back as the early 1980s. These initiatives have had varying degrees of success. The major challenge has been sustainability. The current KMD vision is to have all the data on electronic media as soon as possible. The tasks can be listed as sorting, arranging, inventorying, repairing, scanning and binding, digitization and archival system. Of all these tasks, digitization has taken off quite well. 20 non-meteorologists data clerks have been hired, on casual terms, within the last 2 years. This has enabled main datasets to be brought to almost up-to-date status (June 2014). This has been slower than expected due to inadequate data entry terminals. The other thrust has been in getting the data digitized at the source by equipping the observatories with computers and GPRS modems such that data is being received on real time into the Climsoft CDMS. In spite of these two success areas, there still remains work on QC and establishing robust infrastructure maintenance and upgrading system. Most of the other tasks are still pending. In summary, the main challenges facing data rescue in the Service, is administrative and budgeting rather than technological or capacity.

Chenghu Sun made a presentation on DARE progress in China Meteorological Administration. During recent years, the large progress in DARE had been made in China Meteorological Administration. In 2013, the guidance of meteorological data rescue in China had been issued, which is the obligatory rule of CMA. Under this guidance, CMA had released a list of rescued datasets including precipitation, surface air temperature, wind, humidity, pressure, soil moisture, soil temperature etc. Daily data from more than 2400 stations for the

time period 1951-present have been digitized and quality controlled too. Hourly data of precipitation, surface air temperature, wind, and humidity have been digitized and QC'd from more than 700 stations from 1961-present. Another important accomplishment of CMA was the implementation of two reanalysis projects, the global atmosphere reanalysis and the East Asian regional atmosphere reanalysis projects, which would cover the period from 1979 to present which is scheduled to be finished within 4 years.

Rob Allan provided an overview on the Atmospheric Circulation over the Reconstruction over the Earth project (ACRE). The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative (<http://www.met-acre.org>) undertakes and facilitates historical global surface terrestrial and marine weather data recovery, imaging and digitisation, feeding these data into the international repositories (International Comprehensive Ocean Atmosphere Data Set [ICOADS] and the International Surface Pressure Databank [ISPD]) responsible for such material, seeing that these repositories provide the best quality and quantity of surface weather observations for assimilation into all reanalyses, and ensuring that reanalyses outputs are freely available and feed seamlessly into the climate science, climate applications, impacts, risks and extremes communities (Allan et al., 2011). ACRE is also expanding to develop an integrated cross-disciplinary focus (climate science melding with social sciences and humanities) on historical reanalyses and weather reconstructions, ensuring that the global historical weather observations and reanalyses outputs are analysed and assessed in a longer historical context, and tailored to the needs of educators, students and the general public.

ACRE is run from the Met Office Hadley Centre, but relies on the continuation of 'grassroots' support from the international weather/climate data community and some funding and in kind support from a core consortium of nine partners: the University of Southern Queensland in Australia; the US National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) and Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado; The National Climatic Data Center (NCDC) of NOAA; the International Environmental Data Rescue Organization (IEDRO); the University of Sussex in the UK, the British Library; the University of Giessen in Germany and the University of Bern in Switzerland. Under ACRE's broad international focus, it has worked to develop various regional data rescue foci, such as in Chile, the Pacific, China, Canada, Meso-America, India, SE Asia, Africa, Arctic, and the Southern Ocean/Antarctica, all of which are at various stages of development. See latest details at:

<http://www.met-acre.org/Home/2014%20ACRE.docx?attredirects=0&d=1>

Mr Allan promoted the publication of more scientific papers around rescued or to-be-rescued data and mentioned the Geoscience Data Journal, for which he serves as the Editor-in-Chief. He also recommended collaboration with other communities on data rescue, such as National Council of Archives, universities etc.

Issues discussed

- 1) DARE efforts in developing countries usually pose sustainability issues. It is crucial to consider how to manage the risk of losing the momentum when DARE projects come to an end.
- 2) AWS and sustainability, maintenance and calibration which rely essentially on NMHSs despite the fact that the equipment was provided through a donor project. There is a need to provide guidelines and maintenance personnel training to ensure sustainability in recipient countries. If the guidelines are not being followed, an observer needs to be kept at the site if at all possible.
- 3) In discussing the potential of crowdsourcing, the team raised the question of how to enlist the help of individuals and communities with crowdsourcing activities. Examples were given of how individuals and whole communities have benefitted both intellectually and professionally. When recruiting participants for crowdsourcing activities, this positive win-win situation should be shared.
- 4) On Resource Mobilization for DARE, the team emphasized the great need for mobilizing resources for DARE projects and activities. Efforts by the team should help in developing business cases that can be brought to the attention of donors and for fund raising through WMO and GFCS channels. Engaging other potential donors (from private communities) is also something that ET-DARE can consider. Also the team felt the need to raise awareness amongst the Directors to make DARE a priority for their investment.
- 5) DARE should not focus on atmospheric data only, but also on other data, such as marine, sea level, and hydrology data. It should also extend to other type of archives such as strip charts.
- 6) The team feels that character recognition technique such as OCR/ICR is not yet ready for DARE due to difficulties this technology has in adequately reading manuscripts with different styles of handwriting. The use of OCR/ICR in many cases shows that manual intervention is always needed to correct the outcome of OCR. Under certain circumstances OCR could help in digitizing well printed archives.

Recommendations

- DARE projects: Maintain contacts and communication to ensure DARE efforts continue. Link DARE to applications and capacity building.
- AWS: Provide guidelines to countries on AWS. This might be something that ET-DARE liaise on with other WMO expert groups (CBS, CIMO)

- Crowd-sourcing: Develop a paper (e.g. Guidelines on crowd-sourcing), develop a communication flyer, and online expert help.
- Seize the opportunity of the DARE side event at Congress-17 to raise awareness of DARE and present a solid business case on I-DARE.
- Ensure linkage with JCOMM and CHy appropriate Expert Teams on Data Rescue, including for example sea surface temperature and stream flow data.

Specific actions:

- Assess existing crowd-sourcing experience worldwide (Rob Allan)
- Advice on Portugal expertise in using OCR technique (Rob Allan)
- Investigate engaging private sector with DARE (Rick Crouthamel)

4. Review of Terms of Reference.

The Group reviewed its ToR and suggested a small change to make current TOR a) a general statement:

The ET-DARE will engage in activities that lead to more data being rescued, including but not limited to the following Terms of Reference;

- a. Establish and record, through contact with interested parties including data users and data centres, general and specific needs for the rescue of historic observational data and metadata records;
- b. Assess regional needs for data rescue projects and investigate associated synergies across different regions, WMO Programmes and Commissions and other international climate data recovery efforts;
- c. Arrange to implement, populate, and maintain an International Data Rescue web portal (I-DARE) that summarizes key information and provides a gap analysis w.r.t. data rescue activities internationally;
- d. Develop, and amend as necessary, guidelines and best practices for the rescue, inventory, hard-copy and electronic storage, imaging and digitization of climate data records;
- e. Collaborate with groups associated with the delivery of climate services and research around their priorities for data rescue, including relevant workshops;
- f. Explore efforts in crowdsourcing as a data rescue strategy (e.g., <http://oldweather.org> for marine data and citizen science alliance).

5. International Data Rescue Portal (I-DARE)

In the presentation on the International Data Rescue Portal (I-DARE), Peter Siegmund presented the main elements of the recently written White Paper on this portal. Subjects that were outlined are: the history of I-DARE, the connection of I-DARE to WIS and to GFCS, the idea and benefits of I-DARE, items proposed to be displayed on the I-DARE portal, governance of I-DARE, resources, and a proposed timeline for I-DARE. ET-DARE endorsed the white paper and agreed to help in the design and implementation of I-DARE (White Paper cf. Annex 6). The team agreed on the following recommendations:

- 1) Integration of I-DARE into WMO Information System,
- 2) I-DARE should involve land and sea data and other data, and establish a schedule for updating the content,
- 3) Report to ICG-WIGOS,
- 4) Standardize regional web portals,
- 5) Look at the experience on inventory in Australia and elsewhere,
- 6) Develop a standard template for DARE inventory as soon as possible,
- 7) Work on multi-level efforts involving global centres down to national,
- 8) Engage countries in developing the inventory on their own, but provide guidance on how to do it,
- 9) Look at the existing resources in Regional Climate Centres (regional infrastructure for I-DARE).

Actions/Decisions:

- Share the white paper with other communities, ISTI, Marine, Hydrology...
- Agree on setting 5 working groups:
 -
 - o **WG on Data Portal:** Peter Siegmund (Facilitator), Meaghan Flannery, David Muchemi, Chenghu Sun, Nancy Westcott;
 - o **WG on seeking advice from other stakeholders (ISTI, marine ...):** Rob Allan (Facilitator), Rick Crouthamel, Peter Siegmund
 - o **WG on inventory (best practices, forms, guidelines on how the inventory is done):** Rick Crouthamel (Facilitator), Rob Allan, David Muchemi, Jose Carlos; Rob to ask Antonia Valente from Portugal if she is willing to join;
 - o **WG on updating the WMO Guidelines:** Nancy Westcott (Facilitator), David Muchemi, Dona Cuppett, Meaghan Flannery (add a QC section);
 - o **WG on communication pieces for I-DARE portal for Congress:** Rick Crouthamel (Facilitator), Rob Allan, Peter Siegmund, Meaghan Flannery, Dona Cuppett.

6. Key Activities

This item was introduced by two presentations:

i) **Steve Foreman** made a presentation on WIS and its relevancy to Data Rescue in addition to information on the concept of Metadata within WIS (Discovery Metadata) and WIGOS (Station Metadata). The meeting welcomed the presentation and agreed that WIS offers a good infrastructure for supporting I-DARE and provides the framework for data discoverability and sharing information.

ii) **Omar Baddour** presented a proposal for guiding Members in developing and least developed countries in order to supplement their request for DARE project assistance with quantitative information on existing archives that need to be rescued and digitized. ET reviewed the proposal and made few additional input. ET endorsed it as a basis for requesting information on DARE projects at national level (Annex 5).

I-DARE specification and design

The meeting agreed to pursue the development of I-DARE as an immediate activity to be carried out collectively. It proposed an expert meeting (Peter and Meaghan), 8 – 12 December 2014, to devise the I-DARE project plan, building on the I-DARE White Paper and covering the following:

- 1) Expand content of project to define timelines, deliverables, and people/groups involved in the plan, make use of figures similar to those in the presentation by Denis Stuber in the joint ET-DARE/CDMS session on 5 November 2014,
- 2) Establish timeline and milestones to enable presentation of Beta I-DARE portal at WMO Congress May/June 2015,
- 3) Set WG outcomes and timelines within the I-DARE project plan,
- 4) Estimate costs of consultancy IT work for Beta version and thereafter,
- 5) Create statement of benefit to stakeholders and NMS's,
- 6) Integrate the recommendations arising from the working groups investigating stakeholder group feedback and suggestions; and undigitised data inventory solutions in managing regional differences in cataloguing station numbers. These recommendations to be provided by the 2nd week of December,
- 7) Develop communications plan for the PR's promoting the benefit of I-DARE with enhanced climate applications output,
- 8) Standardise layout and front end of I-DARE portal, to be finalised by Feb/Mar 2015 to inform the graphics contractor (Dona's contact) responsible for pamphlets, posters, logo, buttons, style and short video. IEDRO expertise / experience to guide this activity. Budget to be determined during December meeting. WMO can print the publicity materials,
- 9) Conduct a hook up / Skype meeting mid-week during the I-DARE project plan meeting in December to inform and ask feedback on the development of the I-DARE project plan over this week from the overview and comments from the ET DARE team.

The activity of other I-DARE Working Groups

- i) Stakeholder WG: Feedback and suggestions from existing DARE portals to provide advice to the I-DARE WG by 2nd week of December
- ii) Guidelines WG: Explore leveraging the use of existing guidelines for crowdsourcing and QC from existing sites such as ACRE, Old Weather. Timeframe for devising guidelines: after Congress
- iii) Inventory WG: Provide a list of challenges anticipated with inventory upload to the portal to I-DARE WG by 2nd week of December. Explore a data base management system solution that allows for variation of record retrieval. Investigate organising the inventory so that users can determine: What needs to be scanned? What needs to be digitized? What needs to be QC'd? ET DARE members to provide examples (scans) of data forms/autographic charts to Rick.

7. Workplan

Activity	Timeline	Responsibility	Expected deliverable
TOR a): Record general and Specific needs for rescuing Data and Metadata			
Inventory – define needs, request from PR, populate I-DARE	Before Congress, from our countries and NMHSs and known inventories, after Congress from other countries	WG on Inventory	Inventory
Other data not normally acquired by NMHSs	After Congress, for next ET-DARE meeting	WG on inventory	Inventory
TOR b): Regional needs for DARE projects			
Liaise with RA WG on climate WMO send focal point of the RA WGs	Ongoing	All ET members Copy to ET	
Link with MEDARE SACA&D WACA&DARE LACA&D MNM Jamaica workshop, University of West Indies and NMHS of Jamaica COSPPac (15 PICS)		David for RA-1 Peter for RA-2 Meaghan (RA5) Chenghu (RA2) David & Peter Jose Carlos Peter Nancy and Teddy Meaghan	
Link to ACRE regional activities	Ongoing	Rob Allan	Feedback at the next ET meetings

TOR c): I-DARE			
Expert Meeting	Dec 2014	Peter and Meaghan	Specifications of IDARE and business case
Prototype	March 2015	Peter	Prototype
Side Event Congress Investigate who is coming (Tom, William...)	June 2015	Peter (PR of KNMI); Omar to investigate with Tom	Side event
Communication material: Peter to send information for the design of I-DARE communication pieces; WMO printing	February 2015 May 1, 2015	Rick lead Team to review (Peter) Rob helps on Video	
A) Stakeholder feedback on what they need to use it, to get commitment, B) feedback on the designed portal implementation	A) 1 December 2014, B) begin after Congress	WG I-DARE – Rob Allan	
TOR d): Update WMO Guidelines on DARE			
Review guidelines in existence on DARE, identify what should be in guidelines and update guidelines	2018	WG on guidelines, Facilitator: Nancy	Revised guideline on DARE, vetted through WMO
Create Wiki		David Moderator	Wiki
WMO set a wiki page	December (Peer)		
Gather information on technologies, i.e. experts in the various technologies	Ongoing	WG on guidelines	Information on Wiki then I-DARE
TOR e) Collaboration with others: Climate Services and Research			
Identify stakeholders	Ongoing	All Omar (for GFCS community) Rick (NOAA office international activities , USGS, NWS)	List of Stakeholders
Link with Association of State Climatologists	Ongoing	Nancy	Feedback to next meetings
Link with WMO RCCs	Ongoing	David (RA1) Chenghu (RA2) Jose Carlos (RA3) Meaghan with BMKG and NIWA	Feedback to next meetings

List of WMO-RCC focal points Link NOAA-RCCs	December	Peter (RA6) Peer Nancy	
Link with WCRP	Ongoing	Peter	Feedback to next meetings
Link to CAgM, CHy	Ongoing	William Wright	
TOR f): Explore crowdsourcing as a DARE strategy			
Seek information from OldWeather; from IEDRO, from Slonosky on terrestrial data, charts, motivation for land data (games, prizes, interested groups and seniors, school kids, climatologists)	Begin now	Nancy Westcott, Rob Allan, Rick C.	Feedback to next meetings
Build plans for new crowdsourcing	After inventory and new DARE guidelines	Same	New crowdsourcing projects

8. Any other business

Specific actions:

- Seek the concurrence of KNMI to help on I-DARE design and development, prototype for Congress-17 (Peter)
- Ask PR of Netherlands to lead the side event on I-DARE (Peter)
- Liaise with other WMO Techn. Commissions on DARE (William)
- Establish web site (WIKI) for ET to share information and practices (WMO, David)
- Develop communication material for the side event: flyer, poster, etc. (IEDRO to assist)

9. Recommendations

- 1- Promote and communicate socio-economic value of DARE
- 2- Work closely with ET-CDMS making sure QC and DARE activities are developed jointly
- 3- Explore joint project with ITU on climate change, adaptation, mitigation with the focus on improving countries connectivity, training, and resourcing
- 4- Reach out to WHO, UNICEF, UNESCO, FAO
- 5- Prospect the value of UN Big Data initiative to the ET work and seek potential contribution from WMO and GFCS perspective <http://www.unglobalpulse.org/about-new>.
- 6- Seek best strategy for involving PRs and climate Data Managers to consider DARE more closely
- 7- Encourage Managers of the NMHS to work on DARE
- 8- Consider other old media, i.e. microfiches, magnetic tapes. Locate old card and electronic readers and see how to make use of them. Develop migration strategy for storage media.

- 9- Develop marketing strategy (benefit of DARE) with the Met Services and Stakeholders; and investigate investment opportunity with WB through WMO Public Relations.
- 10- Concentrate attention and effort on Education & Training, capacity building and sustainability.
- 11- Involve young scientists with DARE work, for example, using social media, competitions
- 12- Establish a pool of DARE experts to help conduct field DARE missions,
- 13- Promote DARE opportunities to volunteers, e.g., senior colleges, universities and high schools, local historical societies, community service clubs, NGOs.

Next meeting: Possibly tie with ACRE meeting in Santiago in October 2015.

List of Annexes:

- 1 Agenda**
- 2 List of participants**
- 3 Agenda of the joint ET CDMS – ET DARE session**
- 4 Draft records of the joint ET CDMS – ET DARE session**
- 5 Guidance for submitting DARE proposals**
- 6 I-DARE White Paper**