

Data rescue task team call

Dec 5th at 13Z (14 CET, 08 EST)

Present on call: Peter Thorne (PT), Jared Rennie (JR), Steven Worley, Rob Allan, Manola Brunet, Jay Lawrimore, Stefan Bronimann, Hermann Maechel

Apologies in advance: Rick Crouthamel, Juerg Luterbacher (called away at short notice)

Based upon unanimous agreement task team members saw value in continuing this task team. As such a number of actions were agreed.

ACTION: Peter to revise and circulate ToR for comment and feedback.

ACTION: Jared and Rob to ensure that ACRE facilitated data are pulled through to databank.

ACTION: Jared to follow up with Manola on EURO4M data.

ACTION: Jared provide to Manola Mexican manifest in stage 2 + stage 3 to ensure against redundant efforts.

ACTION: Jay to ensure databank working group to discuss IMMA on their next call. A good candidate for the ISTI discussion would be Scott Woodruff. Scott was on the ad hoc ACRE group on this.

ACTION: Jay to nominate someone from NCDC (pref. involved in databank) to sit on Rob's ad hoc group on creating a master data rescue activities directory.

1. Updates on ISTI activities and databank - Peter / Jay

a. Overall programmatic stuff - Peter

ISTI continues to move forwards. Major things in recent months:

1. A letter was sent from WMO asking NMSs to provide feedback and additional data to the databank. Anecdotally this may have shaken out at least some new data holdings.
2. We got a SAMSI summer program grant accepted so next summer there will be an event at Boulder to try to engage people in developing new and novel homogenization approaches.
3. The benchmarking is starting to progress quite fast.

b. Databank - Jay

The databank paper has been accepted by Geosciences Data Journal. This was to trigger the release of version 1. At the eleventh hour NCDC management have mandated an Operational Readiness Review. This review is in progress and should be finished on the timescale of 1-4 months at which point the first version release will occur. In the meantime work is ongoing to enable NRT updates and periodic period of record updates. The databank working group is working on many additional things viz:

1. Collecting parallel measurements (side by side instrumentation)
2. Metadata
3. Additional data sources
4. Improving daily holdings

The databank WG recognizes the value the task team has brought to databank development but does not wish to mandate task team continuation. If the task team does not continue then thought needs to be given as to what steps can be taken to ensure continued collaboration with DARE activities in an effective manner.

2. Discussion of continuation of the data rescue task team - Peter

The task team has been useful and certainly has helped to pull through data to the databank version 1 release which we otherwise would not have. However, over the last 18 months it has been hard to gain traction. This is partly because I was sidelined by IPCC and something had to give for a 12 month period. It is a pity that it was this important task.

The databank working group and steering committee have both discussed the issue, and both have effectively deferred to this group as to whether they see value in continuing to operate as a stand alone task team activity.

If the membership wishes to disband then the plan would be to fold in a better coupling between ISTI and data rescue activities through 1. expanding the databank ToR and membership a little and 2. trying to gain greater involvement of those working on the databank on other international DARE activities.

Following these opening remarks those on the call were canvassed for an opinion. Members on the call unanimously saw value in continuing. They also suggested that remit be clarified.

We need to broaden scope - to other elements and metadata.

What can a few volunteers achieve? What is the sustainability? Is task to exchange information? Or is it something more? ACRE does more. Need for sure registry and coordination. Is ISTI the right forum? What is the focus and scope?

ACTION: Peter to revise and circulate ToR for comment and feedback.

3. Updates on data rescue activities and opportunities - All

Justus Liebig University of Giessen, Germany:

There are different ongoing activities of digitising subdaily meteorological data: Below is an incomplete list of stations that are currently digitised and the years mentioned that will be available in January 2014.

St. Catharinenburg (Russia); 1837-1844
Zlatoouste (Russia); 1837-1880
St. Petersburg (Russia); 1837-1845
Harbour Office (Hong Kong, China); 1861-1885
Valencia (Spain): 1790- early 20th century
Antigua (Caribbean): 1893-1900; 1922-1930
George town (Guyana); 1887-1950
Moreui, Lokoja, Yola and Zungeru (all northern Nigeria): 1907-1910
Mopea Zambesi River (Mozambique, East Africa); 1901-1904
Dar es Salaam station (Tanzania); 1926-1940
Seychelles; 1912-1919
Ships 'Gauss' and 'Kerguelen'; 1902-1903
Halifax Murrison (Canada): 1835-1841; 1846-1850
St John's Times (Canada); 1832-1837
Kingston, Fort Simpson and New Westminster (Vancouver) (Canada), waiting for scans and periods

All those stations will be completed with additional years within 2014. There is a particular effort to work on a variety of old Russian stations with some starting in 1837 and most ending in 1880. Peter, Jared and Rob have a .pdf file with a compilation of station names and length of series. We will start with the three longest ones (Zlatoouste, St. Catharinenburg and St. Petersburg) going back to 1837 followed by many other stations. We will digitise the subdaily pressure, temperature and precipitation data.

This will keep us busy in 2014.

Steve Worley: There is some very nice work going on at JLUG, Germany, and ICOADS should get the ships Gauss and Kerguelen (obviously Southern Ocean) data on the list for future inclusion.

Rob Allan: Note that the great bulk of this is coming from ACRE activities in scanning the material, which is then sent onto JLUG for digitisation. JLUG also fund my partner Gail Kelly to digitise various historical terrestrial and marine weather observations under ACRE

RA: Will also ensure these data are pulled through to the databank.

ACTION: Jared and Rob to ensure that ACRE facilitated data are pulled through to databank.

SW: Would like to ensure that ship data are pulled through. Would like to be engaged. Coordinate with Rob, Scott, and efforts in the UK that are handling many of the ship based ACRE pieces.

MB: EURO4M daily Tx, Tn, precip and SLP from around 60-70 Mediterranean stations. These observations are almost ready to be provided. To be uploaded soon. Will be available to ISTI databank. Includes metadata. Can facilitate the upload.

ACTION: Jared to follow up with Manola on EURO4M data.

MB: This year will start new ERA project on synoptic scale data. 2 years envisaged til available. Scope is to be exactly defined with ECMWF.

RA: Should have some North African stations from UK Daily Weather Report digitised data to Manola soon.

SB: Asked Antonia to share data. Will be more data soon from Chile. Limited surface data – focus has been on UA data rescue but will share what we have.

SB: Started to gather data on the year without a summer. Data uploaded to ISPD. Jared should be able to get from there?

HM: No new data.

MB: For Europe ERA-CLIM2 will provide more data.

MB: Some NMSs are trying to set up data rescue activities. Mexican pre-1950. Scan and then later digitize. Mexico will be open access data.

RA: Check that CDMP have not done this already?

ACTION: Jared provide to Manola Mexican manifest in stage 2 + stage 3 to ensure against redundancy of efforts.

4. Discussion of data format, data rescue documentation and data rationalization issues raised from ACRE meeting in Lisbon

The text below is purloined from an email thread. Its only partial quotes

Scott Woodruff:

On the observational format side, it does seem like there could be some attractive advantages (e.g. facilitating greater interoperability between marine and land data--including shared development software and documentation) in transitioning basic (i.e. sub-monthly res.) land station data into the flexible and extensible IMMA format, which is used widely for ICOADS.

As Gil mentioned, the IMMA format is already archive-ready. If this approach seems worth exploring further, possibly the emphasis, at least initially or on a "pilot" basis, could be to provide an internationally recognized and robust archival target format for newly available land observations, so as to reliably preserve temperature plus all other reported data elements (including a copy of the original input data).

Earlier this week, Gil and I talked over some of the detailed field changes to the current IMMA format that would appear to be necessary if we were to go down this road. I'm not at all well versed in exactly how the land-station data are currently exchanged internationally, and archived, or regarding what specific synoptic data fields are typically reported over GTS or in delayed-mode (note: also for contemporary versus historical data requirements), so the details probably would need quite a bit of refinement. However, I hope this brief document will provide some useful ideas for further discussion of the pros and cons of this potential formatting approach (including possibly some variations, as noted, on the specifically outlined solution).

Albert Klein Tank:

My recollection of the discussion in Lisbon is that the need was in particular for an online database / overview showing the metadata of the data that are being rescued and digitized in different projects / activities and where these data can be found/downloaded. So, more or less an extended and updated online version of the excel spreadsheet you maintain. That doesn't mean that I'm against developing more integrated databases, but that may be too ambitious for now, and it doesn't answer Dick's question (or a similar question Gil may ask) about what additional data for reanalysis he can expect in the next few months or years.

Issue of curation is a large issue. Who looks after a master listing for the long term?

WMO ET on data rescue?

RA will follow up with contact in OPACE1 on suitability of the WMO ET.

At least two threads: Data format, data rescue metadata and master listing to ensure we do it smartly and prioritize correctly.

RA: Will be pilot on IMMA type format next year between UKMO and Juerg's group.

ACTION: Jay to ensure databank working group to discuss IMMA on their next call. A good candidate for the ISTI discussion would be Scott Woodruff. Scott was on the ad hoc ACRE group on this

ACTION: Jay to nominate someone from NCDC (pref. involved in databank) to sit on Rob's ad hoc group on creating a master data rescue activities directory.

5. Any Other Business

There will be an ISTI all hands call in January. Encourage all to take part. Here our revised ToR will be discussed and adopted.

We should have a next call in March

ACRE: 25th-27th August in Toronto. At Environment Canada