

Data rescue task team call 11/15

8am EST, 13GMT/Z, 14CET etc.

Dial in details sent by email

Present on call: Peter Thorne (PT), Jared Rennie (JR), Rob Allan (RA), Stefan Bronnimann (SB), Jay Lawrimore (JL), Rick Crouthamel (RC), Hermann Machel (HM), Juerg Luterbacher (JL)

Apologies: Rod Hutchinson (RH), Manola Brunet

1. Review of actions from last call

ACTION: All to provide any feedback on poster draft by 9/7. PT to redraft based upon feedback and then send back out. Final comment and redraft and send to printers 9/31.

The poster was completed and printed and used at WCRP OSC in Denver in October. It was one of a cluster of four initiative related posters and there was also a talk given in the parallel session on observations. Poster solicited at least one blog posting <http://www.easterbrook.ca/steve/?p=2714>

ACTION: PT to try to push forwards formation of a team of interested parties who might prepare a white paper to document a plan for how data will be stored and used by the climate community more broadly in support of renegotiation of resolution 40.

This was discussed at the steering committee meeting and then over several informal meetings with the president of the Commission for Climatology (CCI). CCI discussed a submission prepared by members of the steering committee databank working group and this task team at their executive committee meeting and will be considering further their position between now and the WMO meeting on GFCS next year.

RC: IEDRO have changed philosophy in past several months and show what the digitized data can do personally for them.

PT: NCDC had offered to create normals, maybe additional

JL: Extremes?

RC: Produced quick program on data rescued from NIger. Just goes back through 20 years, then add 30 years shows change.

PT: Lisa Alexander making climdex software available?

ACTION: PT to contact Lisa Alexander in this regard. And NCDC normals team.

2. Update on over-arching initiative progress - PT

Both working groups have submitted progress reports which will be made public in the near future from the surfacetemperatures.org domain.

Representation at WCRP OSC was overall successful. It yielded several potential data leads which have been followed up by Jared Rennie, Jay Lawrimore and Peter Thorne. Several how to submit data instructions and data request letter hard copies were taken.

We now have recognition from TIES (the international environmental statistics body)

Several additional meetings have been attended by members of the steering committee ... most noticeably GCOS steering committee and the recent ACRE meeting from this task team's perspective.

3. Review of progress on databank effort - Jay Lawrimore

The Databank has now been populated with numerous sources of data at both the daily and monthly timescales during the past year. Daily data are available in Stage 1 and Stage 2 formats for the following.

- Australia
- Brazil
- Channel-islands
- ECA/KNMI
- Ecuador
- ISPD (International Surface Pressure Data)
- IPY
- Swiss
- Sydney
- Tunisia-Morocco
- Japan
- Mexico
- Pitcairnisland
- SACA/KNMI
- Spain

- Uruguay-INIA
- US Forts
- Vietnam

Monthly summary data are available in Stage 1 and Stage 2 formats for the following.

- Antarctica-SCAR-reader
- Antarctica-South Pole
- Arctic
- Australia
- Canada
- Central Asia
- Colonial Era Archive
- East Africa
- GHCN-Monthly version 2
- GHCN-M version 2 source data
- HadCRU version 3
- Histaip
- Japan
- ECA/KNMI
- Russia
- UK Met Office historical
- World Weather Records

Each source dataset is held in its own subdirectory and a README file accompanies the data in the Stage 1 directory, providing basic information on the data provider and the data. An INVENTORY file is included in each source subdirectory within the Stage 2 directory structure. This file provides a list of each station for the particular source and other metadata including station id, name, location, elevation, and first and last year of data.

A limited amount of Stage 3 and Stage 4 data has been added to the Databank on the Daily side - from the Global Historical Climatology Network-Daily dataset - contains more than 25,000 stations with daily maximum and minimum temperature. This is also available for conversion to monthly summaries and future use on the Monthly side of the databank.

Work has now begun on developing methods of data merging and a process for developing and launching the first version of the merged Stage 3 data. The working group has as its goal the launch of version 1 of the global databank in April 2012. The development is being tracked through a wiki at http://editthis.info/intl_surface_temp_initiative/Main_Page. The current thought is to use an explicitly bayesian framework to decide on duplicate assignments and merge priorities. This will be iterative, merging higher priority data decks first and then progressively lower priority decks later. Prioritization has yet to be finalized.

To aid the collection of new source data - a Data Submission Guidance document is now available. This provides instructions for data providers - a description of the process for providing raw native digitized (Stage 1) data to the databank and encourages data providers to follow basic guidelines. A cover letter to accompany the guidance is also available and provides a brief overview of the International Surface Temperature Initiative and the role of the global databank within the wider context of the grand initiative. The cover letter and guidance document are available on the databank website: <http://www.surface temperatures.org/databank>.

A certificate of appreciation also is under development. This will convey to National Meteorological and Hydrological Services the gratitude of WMO's Commission for Climatology, the Global Climate Observing System, and the World Climate Research Programme for their support of the International Surface Temperature Initiative and broader efforts to meet 21st century needs for climate information.

Objectives for the coming year:

- 1) Continue to add sources of Daily and Monthly timescale data to the Databank. Work with DWG members and others in identifying and collecting readily available sources of digital data.
- 2) Build upon Data Rescue activities and leverage crowd sourcing efforts to begin volunteer digitization of land surface records.
- 3) Complete position paper on version control and provenance available for public comment.

- 4) Develop an approved methodology for merging sources of data to create a monthly Stage 3 data product. Include a hierarchy of source data from which to build the merged dataset.
- 5) Launch version 1 of the Databank for monthly timescale data in April 2012, making all data, processes, and software freely available and accessible.
- 6) Complete and submit journal article describing version 1 of the Databank and its underlying principles.
- 7) Work with the Benchmarking and Assessment working group to expand opportunities for incorporating Stage 4 and 5 data into the databank.

4. Update on data rescue pull through to databank

Thanks to Juerg Luterbacher for data submission to the databank. We are, at time of typing, in the process of adding this to the databank.

We have in principal agreement to include data from the Canadian data rescue effort.

Other additional submissions welcome. The databank will be a living, breathing, all singing and dancing thing so we can add to holdings you or others submit later. At this stage to aid our decision as to merging order it is useful to have **any** data submissions from likely sources so please submit something ASAP to aid the merge program development. Don't let the perfect be the enemy of the good.

RA: GHCN-M provenance is there a hard copy of original?

JL: Maybe but hard to be sure, would be a lot of work. Back in 1990s there wasn't the focus on provenance, more the final copy.

SB: Do you need submission regularly timetabled?

PT: No, just as and when, keep as complete as possible.

JR: The databank is structured so that if new data comes in within that source, we can update that source and re-run the code without major problems. I would highly encourage updates as soon as possible

5. Update on funding hopes.

JL and PT have resubmitted a funding bid for a crowdsourcing solution on the land data through the Citizen Science Alliance and CICS-NC which, if funded, would start up in FY12 (the coming 10 months)

PT chased google.org around a quasi-bid put in early in the year around imaging data in the NCDC basement and / or helping to digitize the holdings in the NOAA Foreign Data Library. I am yet to hear back anything concrete beyond that it was still under consideration.

IEDRO update on google funding?

RC: We wish to send in a large proposal to google. One of volunteers went to work in New York for google. Were notified amongst 8 finalists. Top prize is \$10,000, 3 of \$5000. Find out Thursday. May be an opening to a larger bid. Would look to work with the Citizen Science Alliance.

JL: Weatherwizard being used?

RC: Predates the oldweather.org effort. But only an idea.

RC: Working on strip chart digitization to digitize precip measurements using an automated program. KNMI also working and have similar. Thermograms or barograms are next on the list. Do we have data digitized from thermograms pre-existing?

JL: Unaware of any.

PT: Would want daily max and min and hourly.

6. Report from ACRE meeting - Rob Allan

4th ACRE workshop was held in De Bilt at KNMI. 50-60 people. Report is up on the ACRE website. Will send general email. Still pushing a number of efforts. Inventory updates and data recovery efforts. Most difficulty is keeping funding for people to be engaged on data rescue, recovery and inventory work. Looking at next meeting be in Chile.

7. General updates on activities from task team members

PT understands that IEDRO were successful in their recent meeting with ACMAD in at least trying to image the microfiche and that Albert Mhanda (who sits on the steering committee and the databank working group) will act as ACMAD lead on this effort.

RC: We met with Exec Sec of ACMAD and had a very good meeting. ACMAD will let IEDRO borrow the microfiche and to get the imaging done and onto a more stable medium. Contractor quotes \$55K. 20-30 million images. So, need to find funding to get it done. Purchased low volume microfiche imager. Would take many years to do so would rather do it through the contractor. Sent proposal to ACMAD looking to input. Funding proposal covers digital imaging and some funding for professional / crowdsourced keying.

RA: Table of contents: Synoptic records, French colonies have long records, but British are much more limited. Some may be redundant. British may have more in colonial archives. Very valuable, but may be different value in other archives.

RC: Idea is to digitize non-redundant data first. Total sites is c.980 and back to 1910.

RA: Senegal and Mauritius back to 1850.

PT and Jared reddie followed up an OSC lead to Jun Matsumoto (Tokyo Metropolitan University) who is undertaking some digitization for Japan and SE Asia. We could try to find out more if of interest.

RA: Will send some contact info of other work in the area.

SB: Just sent around ERA-CLIM inventory. Not too much surface data.

RA: Block of data from Portugese effort with South China Sea. Twice daily temperature data may be able to be added to the ERA-CLIM effort. 10am and 4pm readings

RC: Have some volunteers who have talked into GIS oriented page on the IEDRO website. Loaded onto the system is a listing of stations for which NCDC databases contain data. Wish to add to this list of data from WMO-A to see what the WMO station list looks like in comparison. Will give a guide to where data doesn't yet exist. Want to build interactive tools.

JL: Sounds like quite an effort.

PT: Idea is that the databank hopefully reduces the need to mine hundreds of data sources. Databank should be repository for all data. Please let Jared Rennie know where any data not in the databank is. (jared.rennie@noaa.gov)

RA: ACRE is trying to get Indian MoU on data recovery.

PT; Jayashree Redvakeer has indicated she is working on Indian data provision to the databank.

RC: Some IEDRO volunteers asked is there any consensus on where there is no data available where people need?

JL: We have maps up on the databank area by decade. But this is only the digital data.
<ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage2/MAPS/>
<ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/daily/stage2/MAPS/>

8. AOB

Propose that next call occur in Feb of 2012

Rod Hutchinson will be an apology. My heart attack problem not resolved and it will be midnight local time and I don't think my wife is keen on me staying up to late.

Digitising funds mentioned last hook-up will be made available though only AUD\$400K. Vanuatu will undertake digitising of uncompleted work begun two years ago plus rainfall only sites funded by Australia. I'll ask Vanuatu if they are happy to put their records up on Temp site. Australia will also fund historical digitising of Fiji records.

Most SW Pacific nations have a modest, but reliable CDMS, we call CliDE.

My abstract for 10ICSHMO in Noumea is accepted for oral presentation. My use some material from this group including poster.

RA: Will anyone here be attending that meeting?

I haven't done much since heart attack, but will focus on getting database in SW Pacific up to scratch and spend some energy garnering historical records for you.

Meaghan Flannery will attend WMO TT-DARE meeting, on my behalf, in Geneva at end of month.