

Data rescue task team

Present: Peter Thorne (PT), Rob Allan (RA), Tom Ross (TR), Stefan Bronnimann (SB), Rick Crouthamel (RC), Manola Brunet (MB), Jay Lawrimore (JL)

Apologies: Rod Hutchinson, William Angel

1. Introductions.
2. Databank progress thus far.

First call a couple of months ago. Data provenance sub-team has met and discussed methodologies on version control.
Putting together a test dataset example of stage 0 to stage 3 data. Image -> native -> common format -> merged. Using US, Mexico, Vietnam stations.
Discussions with Omar Baddour on improving WWR from decadal to annual updates with headway underway.
GHCN-D and M database to be posted shortly through WDC-A website back to stage 2.

RA: Talk to Yin at NCDC who has been given ACRE data.

JL: Yes, looking to incorporate.

TR: Chile long term data set also of interest. JL: Noted.

3. Purpose of group

PT: Looking to make sure that ongoing efforts to rescue data are fed in effectively to the databank. Collection of digitized data and crowdsourcing are in bounds.

SB: For his crowdsourcing not much interest. Few thousand numbers per day. Data-rescue-at-home.org

RA: oldweather.org is quite successful thus far. 7000 people engaged.

RC: oldweather.org got a lot of press stateside.

RA: Try to make it as interesting as you can.

PT: NCDC are interested at director level

JL: Quality of oldweather.org output?

RA: Pretty good quality material through this process.

TR: NOAA Foreign data library is the key start point.

RA: oldweather.org is being written up quasi-real-time to keep engagement.

SB: ERA-CLIM is looking to use UA records from the NOAA int'l library fields. Important to know what is going on already to avoid duplication.

JL: Group is really important. So much data has been imaged and so little pulled through to the databanks. May be missing some people?

MB: Yes, really important to have new data. Avoid redundancies by inventorying current and past projects on data rescue and ensure we know what is happening. Suggest share our current knowledge as a start.

SB: Contact with other NMSs etc. is important to make its importance known. Group should be about coordination.

RA: Feedback to NMSs what we have actually digitized and link it in, everyone benefits.

RC: Group within IEDRO to find reasonably priced software to produce VAPs to give back. Show value that is added. Otherwise some think main benefit is just getting rid of useless piles of paper. Show the value.

JL: Someone from zooniverse on group?

SB: German data rescue. Will provide a name. Perhaps more important that we link to them.

RA: KNMI. Albert Klein Tank or KNMI colleagues? European data sector.

MB: Meteo-France input.

TR: Connections through WMO to get support yet? JL: Good idea. Regions letters. Will consider this.

RC: IEDRO will be happy to put anything useful on website.

RA: Same for ACRE.

SB: Japanese suggestion.

PT: Get to 12-15 for greater geographical spread.

ACTION: All suggest names to PT within next fortnight.

RA: ACRE has regional foci points.

4. Mode of operation

PT: Web based and by teleconference in initial phase.

RA: Sub-meeting at next ACRE meeting in Toulouse in France. Will be second half of year? RA to update group when this meeting clearer

PT: two monthly meetings, every 8 weeks. No objections raised. Agreed to arrange via doodle once expanded team finalized.

ACTION: PT to set up website area on surfacetemperatures.org domain.

5. Avoiding redundancy

RA: If anyone has a real interest in a given region / location, we can check archives at UKMO.

JL: Daily, Africa and S. America are our main interest right now.

RA: Look at BADC website to see what has already been done.

6. Immediate term aims

PT: Formalize this team

JL: Inventory of what data has recently been updated on daily and monthly resolution?

ACTION: All send what they are aware of. PT then updates surfacetemperatures.org

SB: Send out questionnaires on a regular basis?

RC: WMO had a desire to show project database, nothing done in that area by WMO.

7. AOB

MB: For TR we have downloaded lots of books from CDMP. Starting to digitize 40 stations from N. Africa and Middle East. Will be made available. Would like to make aware which variables and stations.

TR: Great to see an application. Need to know this is already being done.

RA: A lot of stuff for Egypt and Sudan could be tabulated and OCRed. Makes crowdsourcing very interesting / challenging.

PT: OCR not necessarily going to work.

JL: If we can get zooniverse to NCDC that would be great.