

ISTI steering committee call 3/1/13 at 8EST, 13Z, 14CET etc.

Dial in details circulated by email in advance

Present on call: Peter Thorne, Jared Rennie, Xiaolan Wang, Jay Lawrimore, Antonio Possolo, Michael de Podesta, Blair Trewin

Apologies in advance: Albert Mhanda, Tom Peterson, Rob Allan

**Next call will be 3/19 @8EDT (12Z)**

**Summary of new actions or actions stayed over from previous calls**

**ACTION:** Andrea will send Meteomet timeseries by email to Kate and Peter. These are the series that will be used in MeteoMet for possible inclusion of instrumental uncertainties.

**ACTION:** Jared Rennie to host the merge process metadata on ftp and make its availability known through the initiative blog and databank ftp pages.

**ACTION:** Kate to put Jared in contact with appropriate ECMWF people to scope potential to interact the databank with reanalysis feedback files.

**ACTION:** Jared Rennie and Xiaolan Wang to follow up on getting the 400 Chinese daily resolution stations' data into the databank

[It was clarified that these are homogenized records and that we would therefore prioritize them accordingly]

**ACTION:** Jared Rennie and Xiaolan Wang to follow up on getting the 400 Chinese daily resolution stations' data into the databank

**ACTION:** Kate Willett to engage with benchmarking working group and provide update to steering committee on what is planned and a new timeline to release of the first version of the benchmarks. By no later than end of Jan 2013.

**ACTION:** Peter Thorne to convert the document of potential data product creators provided by SC members into a spreadsheet on googledocs which everyone can access and members of SC / WG to volunteer to approach the relevant individuals to encourage their participation in product generation from the databank and in the benchmarking exercise.

**ACTION:** Peter Thorne will liaise with GCOS secretariat and AOPC to ascertain how most appropriately to raise a request for regional national focal points to review the databank release and help us to start to improve the provenance of what we have.

**NEW ACTION:** Peter to follow up with Andrea on Michael presenting an update at Tempmeko.

**NEW ACTION:** Peter Thorne to draft initial progress report and circulate for review and comment.

## **1. Review of actions from last steering committee call**

**ACTION:** Rob to contact Jared about journal submission of databank method paper.

*Done*

**ACTION:** Jay to check and confirm number of days going into the monthlies where we have calculated them is in the flags.

*Will be added as Data Provenance Flag in Stage 2*

**ACTION:** Kate to organize benchmarking WG call. By mid-October.

*Superseded*

**ACTION:** Andrea will send Meteomet timeseries by email to Kate and Peter. These are the series that will be used in MeteoMet for possible inclusion of instrumental uncertainties.

*Pending*

**ACTION:** REC to circulate preliminary list of potential target organizations and individuals to take up the challenge of creating data products from the databank for comment when complete

*Done*

## **2. Review of actions from December joint call with the two working groups**

**ACTION:** Jared Rennie to host the merge process metadata on ftp and make its availability known through the initiative blog and databank ftp pages.

*Ready to go. Will do this week.*

**ACTION:** Kate to put Jared in contact with appropriate ECMWF people to scope potential to interact the databank with reanalysis feedback files.

*Pending*

**ACTION:** Feedback on paper to Jared jared.rennie@noaa.gov by 12/17 @23Z. Jared to circulate revised draft to co-authors by 21st.

*Good feedback was received but given the issues raised in the databank group report the paper recirculation is stayed until we have locked down these changes and incorporated into the paper draft to avoid spamming co-authors.*

**ACTION:** Jared Rennie and Xiaolan Wang to follow up on getting the 400 Chinese daily resolution stations' data into the databank

*Have followed up. Xiaolan does not have the data yet, but will let me know once received*

It turned out that they can contribute the homogenized version of the data for now and they need to go through the formal approval process to provide us with the version 0 data.

Homogenized data is fine for now, will be placed lower in source decks. Long term we would prefer original/raw data

**ACTION:** Jay Lawrimore and Kate Willett to work with their groups to facilitate the populating of a daily resolution parallel measurements database for understanding daily inhomogeneity properties

*Blair will help. No action to date.*

**ACTION:** Kate Willett to engage with benchmarking working group and provide update to steering committee on what is planned and a new timeline to release of the first version of the benchmarks. By no later than end of Jan 2013.

*Pending*

**ACTION:** Peter Thorne to convert the document of potential data product creators provided by SC members into a spreadsheet on googledocs which everyone can access and members of SC / WG to volunteer to approach the relevant individuals to encourage their participation in product generation from the databank and in the benchmarking exercise.

*Pending*

**ACTION:** Peter Thorne will liaise with GCOS secretariat and AOPC to ascertain how most appropriately to raise a request for regional national focal points to review the databank release and help us to start to improve the provenance of what we have.

*I raised this with Chair of GCOS Steering Committee / AOPC but it has gone a bit quiet. Will raise again when AOPC attendance invite appears.*

**ACTION:** Peter Thorne and Andrea Merlone to discuss Tempmeko presence for ISTI.

*Pending*

Michael could present an update for us if we wish. PT to follow up offline.

**NEW ACTION:** Peter to follow up with Andrea on Michael presenting an update at Tempmeko.

*There was a feeling that this larger group call was a useful way to engage the broader ISTI community. Should we have these every six months as a way to engage the broader group?*

Yes. We should try to do that.

### **3. Databank update - Jay**

The beta 2 phase of release has yielded some additional issues turned up by climate blogger Nick Stokes whereby stations were still being duplicated from different sources when they arose from distinct source decks with slightly different coordinates and sometimes very different data (where decks had been qc'ed / homogenized. So, the algorithm was doing what it should but the metadata / data were confounding. A further logic gate was added whereby if metadata and data were in substantive disagreement (close metadata match / poor data match) then the station was withheld. This removed 600ish obvious duplicates globally and has been implemented.

Construction of a blacklist facility has required substantive extra work on ascertaining potential issues. Several approaches have been undertaken:

*Intra-station consistency:*

1. Running through NCDC's pairwise homogenization algorithm and looking for evidence of very big breaks. This found very little. And what was found was plausibly real inhomogeneities in the records not blatantly false merges.
2. Running F-test through a ten-year sample filter for gross changes in variance. This yielded a handful of intra-source intra-station cases whereby either portions or the whole record from that source will be black-listed.

*Inter-station consistency:*

Correlation of anomaly first differences was created for 100+ closest neighbours for each station. There were three specific cases of interest:

- i) very nearby stations very highly correlated
- ii) very distant stations very highly correlated
- iii) very near stations with really poor correlation

Classes ii and iii tend to be missing signs on lat and / or lon and can be easily rectified through ascertaining the correct site location and forcing this through blacklist. The last appears to relate to cases where there are geolocation metadata differences or sub-period data differences sufficient to unique the station. We are in the process of going through these by hand to augment the blacklist file appropriately. This is the largest population of just over 200 stations. It tends to be dominated by Pacific island stations that appear to have reported under myriad names through time. This was also an issue PT came up against in constructing HadISD duplicate set so isn't surprising.

Undoubtedly there will remain some issues whenever we decide to release. The databank will evolve as it becomes used and we gain user feedback that highlights issues we have not considered. But once the

blacklist is constructed and tested we now feel confident it is ready for official version 1 release. We envisage databank release and paper submission by mid-Feb 2013.

#### **4. Benchmarking update - Kate**

Note that the update report is still pending - we need this prior to creating the steering committee annual report.

Kate not present on call. Nil update.

#### **5. Review of where we stand against the Implementation Plan**

As a first step towards writing the report we should consider where we are against the Implementation Plan and whether this Plan is any longer a useful document. The IP is available at: [http://www.surface temperatures.org/steering-committee/Surface\\_Temperature\\_Initiative\\_Implementation\\_Plan\\_version1\\_release.pdf?attredirects=0](http://www.surface temperatures.org/steering-committee/Surface_Temperature_Initiative_Implementation_Plan_version1_release.pdf?attredirects=0) Regardless it runs out in July 2013 so would need to be refreshed in the spring. Is this something we want to do?

Jay: Useful to external folks to see what we plan to do.

Blair: Useful plan for us internally and externally. We should update.

Jay: Databank delay has had knock on impacts for benchmarking.

Michael: Need to assess against funded resource (=0)

#### **6. Steps to annual progress report.**

Based upon discussions PT will draft an initial progress report en route to Hobart next week and circulate for comment. Another call is not envisaged - completion should be possible by email iteration. Aim is to have this done by end of Jan.

**NEW ACTION:** Peter Thorne to draft initial progress report and circulate for review and comment.

Last year we got no feedback from CCI, TIES and BIPM which means either a. we are doing just fine and dandy or b. we need to better communicate our findings to our 'sponsors' (no monetary implications!) and better engender feedback from them. Thoughts?

#### **7. AOB**

I (Peter) have submitted a poster abstract to EGU on the Initiative which Victor Venema has agreed to present for us in the homogenization section (I was torn between this and Chris Merchant's but decided we needed the homogenization obsessives as our target audience at this juncture - if Chris wants and rules permit it could be hung at both assuming they don't clash or I could submit a more generic overview poster to his session if somebody were willing to present it). Its in the session of which Xiaolan is convenor (<http://meetingorganizer.copernicus.org/EGU2013/session/11619>) . Sadly I can't attend otherwise I would have suggested oral. I envisage the poster to be about where we are and inviting groups in the session who happen to walk by to come and play in the data homogenization sandpit we have created.

Xiaolan: Please be clear where benchmarking fits in.

Albert Mhanda: arrangements between IEDRO and ACMAD to rescue data on microfiche are now at an advanced stage, as the Director General of ACMAD has now signed the proposal document from IEDRO. I will keep you updated on further progress.

PT: Q. Is there anything we can / should do to help facilitate this process?

JL: Not presently necessarily but we should if asked.

Next call 3/19 @ 8EDT (12Z)