

## **ISTI Databank Working Group call**

6 May 2015 @12Z (14 Europe, 13 UK, 8 EDT)

Expected on call: Jay Lawrimore (JL), Peter Thorne (PT), Jared Rennie, Victor Venema (VV), Steven Worley, David Lister (DL), John Christy (JC), Colin Morice (CM)

Apologies in advance: Waldenio Almeida, Albert Klein-Tank, Matilde Rusticucci, Matt Menne

### **I. Welcome and review action items from the previous Databank WG conference call (Lawrimore)**

JR – check on number of sources being updated monthly.

The sources updated each month as part of the near real time update system are as follows:

GHCN-Daily

Climat Stream from the MetOffice

Climat Stream from NCEI

Unpublished version of Monthly Climatic Data of the World

JR – to complete new post to databank blog regarding the feedback received to date.

Post was written in December of 2014: <http://surfacetemperatures.blogspot.com/2014/12/what-has-changed-since-version-1.html>

JL/MR – Matilde to follow up with INTA regarding routine updates to Argentine stations and Jay to follow up with Matilde to close the loop.

From Matilde: INTA is going towards an automation of the data transmission, (because in the past the transmission was in a monthly interval) by now, they have half of the stations doing it.

When they have all the stations on line, they will send the data to the National Weather Service, and the NWS will go into the GHDN daily (they assume)

Jay has asked Matilde if they can let us know whom at the NWS they will be sending the data to (or will it be available online)?

JL, PT, VV: Discuss Parallel Measurements issue and decide on whether there should be task team under the DWG or if it should be.

First Parallel Observations Science Team (POST) conference call held on 29 April. Report to be provided below.

JL/JR To follow up on 3 sources passed along by David Lister.

From DL:

- There is the news about HISTALP - with a re-homogenization and other changes which should be completed by early 2015 - according to Ingeborg Auer.

- There will be data from Nordic countries by the end of this year - via NORDHOM - according to Erik Engstrom (SMHI)

- Also, there has been an effort to produce climate series for the Pyrenees (monthly series from about 1950) - the contact here is Marc Prohom (Servei Meteorològic de Catalunya, mprohom@meteo.cat ). We have the data so I think they will be public domain.

Have not followed up yet with these three potential sources of data.

Update - 06/05 - HISTALP release has been delayed by up to one year. The last check on NORDHOM availability was a few months ago but there was nothing available at that time.

**Action: JL: contact Marc Prohom regarding climate series for the Pyrenees.**

All – Provide final comments on Annual report by end of week. JL to complete report and submit to steering committee.

Final report submitted.

MM: To follow up and determine next step for a multi-element dataset effort. Will check into possibility for a 2-day meeting to bring together all potential interested parties from temperature, precipitation, snow, pressure, etc. communities.

JL reported for Matt that there is a lot of interest from the community. Discussed at snow workshop in November.

No immediate plans for workshop but still under discussion.

## **II. Status of Monthly Databank and planned near-term activities (Rennie)**

A document was distributed the evening before the call to highlight recent activities. This document was discussed by Jared.

**Action: JR to post summary document to Databank website on [surfacetemperatures.org](http://surfacetemperatures.org).**

Regarding the potential sources for addition (Table 2 of document) need to determine if some of these are already included in ECA&D. If so won't likely include in next merge. Daily can be brought in through ECA&D's inclusion in GHCN-D. Monthly ECA&D already included in databank merge.

**Action: MM and AKT to determine if potential sources for addition to databank merge are already in ECA&D.**

Peter: Do we have any idea of implications of the proposed changes for the longer-term station and spatial completeness counts? In dropping some of the sources do we lose early coverage?

JR: Have not evaluated yet, as latest merge completed yesterday. Will check and compare coverage early in the record.

**Action: JR and others to determine if we lost coverage in the early years (pre-1950) because of our removal of the 4 homog. sources.**

**Action: JR to share analysis with the group before making a final decision.**

Peter: Given the changes in the pdf are we moving to v1.1 or a v2? I guess the question comes down to do the changes require solely a tech note (v1.1) or should they be more fully documented and reviewed in a paper (v2)? I assume further that what hoops would need to be jumped through at NCEI may come into play here but it is likely worth discussing on the call whether this is really v1.1 or v2 and whether in terms of a GHCNv4 relying in part on a tech note opens up to potential unwarranted criticism. A v2 paper if to the same journal could be CONSIDERABLY shorter and reflect solely the updates and their justification? Mind you even labelling v1.1 and writing as a short technical note the GDJ may be possible? Jay of course is on the editorial board so can advise. Bottom line is that we should be comfortable with how we document and version increment this change so just a note here to discuss this aspect.

**Action: JL and others to make decision regarding whether to increment as v1.1.0 or v2.0.0 with an accompanying journal article.**

JC: Some benefit to putting a paper online with a beta release in advance of actual submission. Good suggestions that have been helpful in completing the article.

SW: Makes sense to go with a 2.0 given the changes discussed here.

DL: CLIMAT streams - Differences between the 2 streams?

JL: Additional stations in the UK Stream.

DL: Also some QC made at the UK.

### **III. Status of Daily Databank (GHCN-Daily) and planned activities for remainder of year (Menne)**

As noted in the document circulated in advance of call - about 1200 new stations were added to GHCN-Daily from the Global Summary of the Day (GSOD). These additions allow a direct id match between the CLIMAT WMO number and GHCN-Daily in the monthly merge. It also permits the calculation of monthly mean temperature directly from daily data for all CLIMAT sites in real-time, which will help overcome CLIMAT transmission problems when they occur.

As noted in the Daily databank/GHCN-D plan discussed on the Oct 2014 call this completes 1 of the 3 sources planned for addition in 2015.

Peter: Also, plans for ISD relevant to be covered here?

GHCN-hourly would be built from ISD basis which historically was in a distinct branch and most of which arises from the Air Force. NCEI is initiating meetings with Air Force and ISD personnel to understand ingest and related issues with the goal of developing a plan for developing a new sub-daily dataset that is constructed with a view to vertical coherency and consistency. Multi person year efforts required to achieve this. Following plan development will get NCEI buy-in.

Peter: Robert Dunn of MO has expressed interest and may be able to add to his workplan for some effort to support.

JL: Matt has been talking with Robert regarding a possible visit to NCDC to work on this effort.

#### **IV. Update on activities of the Int'l Surface Temp Initiative Steering Committee and Benchmarking progress (Thorne)**

Progress has been sporadic over (yet another) international relocation by the Chair. The various annual reports etc. were completed and posted along with a new version of the Implementation Plan. There have been some sporadic updates to the blog and the website but not much.

Most effort has been in the background on a possible 5 million euro Irish grant (unknown chances of success) which would look at various citizen scientist / e-learning aspects of the problem. This would involve employing c.25 postdocs and PhDs on ISTI aims. Strands would be:

- i) data rescue by digitization (initial focus on Ireland, then Africa)
- ii) Crowdsourcing HOMER
- iii) Serious gaming using automated algorithms
- iv) Data service

Current partners are Met Eireann, ICHEC (centre for high end computing) and possibly Google and Powerscourt (visitor centre in Wicklow mountains).

I shall look to have NCEI as a partner if possible. The NCEI role shall be documented as receiving rescued data and putting into databank, ISPD, GHCND and ISD as appropriate. It may also involve providing expert input on PHA. I will need a 2-page short CV off of the NCEI PI (Jay, Matt or Jared?) and a short letter of support (from same person or perhaps higher up in NCEI). NCEI, sadly, will not get money out of this but as the role is to receive and integrate data I assume this is already implicit in their core budget?

**Action: JL to initiate a letter of support for PT's Irish grant proposal.**

Other partners are possible but w/o funding outside Ireland (although we can do travel). In particular anyone interested in co-supervision of PhD candidates from an int'l partner would be looked on positively. Drop PT a line.

Benchmarking is progressing slowly. Benchmarks should be available this calendar year. In part need to decide whether the benchmarks should hang off v1 or v1.1 of monthly holdings. Suggest that Jay and Jared take an action to discuss this with Kate.

**Action: JL and JR to follow up with Kate regarding the version of Databank for benchmarks.**

JC: Where are images?

PT: Met Ireland has a storehouse of early Irish data - to be imaged by subcontractor. And NOAA foreign data library with 36 million images from CDMP program but never keyed.

## **V. Update on activities of the Parallel Observations Science Team (POST) and kick-off conference call (Venema, Thorne)**

VV: The Parallel Observations Science Team has been formed and had its first telecon. More members are naturally welcome.

VV: It also has its own ToR and homepage:

[http://www.surface temperatures.org/databank/parallel\\_measurements](http://www.surface temperatures.org/databank/parallel_measurements)

VV: We have (active) member and associate members (people who would like to be informed and have the possibility to give feedback).

VV: The database will be build by working on papers for specific historical transitions in the climate networks.

VV: We aim to collect and anlyse dataset from many climate regions to that we can draw first conclusions on global biases in the raw observations.

\* Enric Aguilar is leading the paper on the transition to AWS for temperature.

\* Petr Stepanek is leading the paper on the transition to AWS for precip.

\* Theo Brandsma is leading the paper on the transition to Stevenson screens.

\* Kate Willett is leading the paper on changes in the measurement of humidity.

\* Jenny Linden is leading the paper on relocations and possible improvements in station siting due to this. With a focus on villages.

More ideas for papers and leads are welcome.

Suggestions for available parallel datasets are welcome. A large part of the datasets know to us is listed here (but should be updated):

<https://ourproject.org/moin/projects/parallel>

Next deadline: Finish basic data processing of the parallel database; Quality control, homogenization [homogeneity assessment] and computation of indices has to be coded. Published for code review Victor Venema, Enric Aguilar, Renate Auchmann: July 2015

PT: To understand Physical impacts of the siting/instrument. Physical basis to justify adjustments - based on a set of parallel measurements.

A good list of parallel datasets has been put together by the team.

VV: To make list of parallel datasets available to the working group members.

<https://drive.google.com/file/d/0B7sJmg1UW9uGakITUUNRVWVldU0/view?usp=sharing>

## **VI. Other Business (All)**

From AKT: Not much new to report from my side other than some significant data updates in the LACA&D system for South America. In particular, for Chile additional temperature series have been added to the web portal recently. See: <http://lacad.ciifen.org/> In case you do not ingest these data on a

regular basis yet (like you do with ECA&D for Europe), it might be worth considering. Also worth noting is the data workshop at ECMWF in late June. You can find more on this workshop here:

<http://www.ecmwf.int/en/copernicus-workshop-climate-observation-requirements>

**Action: JR to review potential data from Chile as recommended by AKT.**

## **VII. Summarize Activities and Next Steps for coming months (All)**

*Action: PT to add Victor to mailing list and to change worley.steven@gmail.com to [worley@ucar.edu](mailto:worley@ucar.edu).*

*Action: VV to make POST list of parallel datasets available to the databank WG members.*

*Action: JL to contact Marc Prohom regarding climate series for the Pyrenees.*

*Action: JR to post summary document to Databank website on [surfacer temperatures.org](http://surfacer temperatures.org).*

*Action: MM and AKT to determine if any potential sources for addition to databank merge are already in ECA&D.*

*Action: JR and others to determine if we lost coverage in the early years (pre-1950) because of our removal of the 4 homog. sources.*

*Action: JR to share analysis with the group before making a final decision.*

*Action: JL and others to make decision regarding whether to increment as v1.1.0 or v2.0.0 with an accompanying journal article.*

*Action: JL to initiate a letter of support for PT's Irish grant proposal.*

*Action: JL and JR to follow up with Kate regarding the version of Databank for benchmarks.*

*Action: JR to review potential data from Chile as recommended by AKT.*

All-hands call potentially at latter part of June. Include a summary of what proposing regarding version 2.