

**Databank Working Group Conference Call**  
**Wednesday, 15 October 2014**  
**1200 – 1315 UTC**

Agenda

- I. Welcome and review action items from the previous Databank WG conference call (Lawrimore)
- II. Status of Monthly Databank and planned near-term activities (Rennie)
- III. Review of Databank Annual report and 2015 Plans (Lawrimore)
- IV. Update on activities of the Int'l Surface Temp Initiative Steering Committee and Benchmarking progress (Thorne)
- V. Daily Data, Support to the Monthly Databank and the need for a new umbrella to support multi-element data sets (Menne, All)
- VI. Other Business (All)
- VII. Summarize Activities and Next Steps for coming months (All)

Jay Lawrimore (JL), Matthew Menne (MM), Jared Rennie (JR), John Christy (JC), David Lister (DL), Peter Thorne (PT), Colin Morice (CM), Victor Venema (VV)

*To help out some members of working group it was agreed that it is OK to move this meeting back 1 hour (later) in the future.*

**I. Welcome and review action items from recent ISTI conference calls (Lawrimore)**

ACTION: CM. To follow up with colleague regarding data opportunities following meeting in Tanzania.

Followed up but no progress yet. Question of what we could get from them. They would want training and using the data. (No info on POR or number of stations.)

ACTION: JR. Spin up plans to release v1 of Databank. Including release of GDJ journal article. Updating website. COMPLETE

ACTION: JR. Rerun merge – 1 more time. Will update sources including GHCN-D. Send to benchmarking group for their global benchmarks. COMPLETE

ACTION: JR. Complete monthly updating system – GHCN-D and CLIMAT streams. First update the beginning of August – data for July and for past 6 months. COMPLETE

ACTION: JL. To send e-mail to WG encouraging WG members to put out notice. COMPLETE

ACTION: MM/JR –To map out what databank sources are not in GHCN-D. Start to incorporate those. COMPLETE

ACTION: JL/MM – short document on long-term plans for daily side of databank. COMPLETE

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

v1 released on June 30. Well received – journal article. Some blog posts. News releases through NERSC, Met Office, NCDC, NCAR.

Monthly update operational - GHCN-D, CLIMAT – append to existing 32,000 stations. – Posted to databank ftp – goes out on the 11<sup>th</sup>.

JC: How many stations coming in.

JL/JR: Unsure – something in neighborhood of 7500.

ACTION: JR – check on number of sources being updated monthly and which sources.

Have received some feedback from users – how are things formatted, questions about the science. Nothing major.

Moving forward – sources working on adding – 9 or 10 sources; have data in hand. Need to convert to stage 2. Chinese and Germany already converted – need to put on ftp.

Continuing to accept new data – cutoff 28 Feb 15 to get new sources in – then work on doing a merge – should have well more than 32,000 then.

PT: can you complete post on blog for what has changed since release and customer feedback.

Jr: did put a post out about the near real-time update system a month ago. And will do new post.

## **III. Review of Databank Annual report and 2015 Plans (Lawrimore)**

Regarding Parallel Measurements –

PT: Spoke with Victor Venema – he suggested bringing parallel measurement effort under ISTI. Working group or Task team under Databank working group.

Victor Venema called in to discuss parallel measurements issue -

VV: Was thinking of independent working group. Simply curious about other people's thoughts regarding advantages and disadvantages.

DL: seems to make sense to go under the databank wg.

CM: Task team under DWG seems most appropriate. Could confuse group if was a separate entity when comes to including parallel measurements, there will be questions about how much and what appropriate to include in databank – many short records – non-standard equipment.

VV: Advantage – brings parallel work to attention of those already involved. –maybe some new sources – gaining more interest.

VV: large component from metadata to understand what type performed.

JL: how much data.

VV: setting up process – think up to 4 datasets – but has list of 50 to 100.

CM: what is POR,

VV: sometimes a few stations, years. But some data from Peru for example has a couple of decades for ~20 stations.

CM: may be some data appropriate for databank.

VV: agrees good to feed into databank as far as it's allowed – some data only allowed to compute indices – not for actual release of data (particularly precip effort in Germany).

PT: Decision is between JL, PT, VV – 2 KEY aspects – have to have parallel m. to be officially recognize by ISTI by formation of a group. 2) need to keep a good line of communication between it and DWG – so either way if it – ex officio on DWG.

PT: Would form as a new group – but would be either a task team or separate working group and only grow DWG by 1 or 2.

VV: or expert team.

Action: JL, PT, VV to continue discussions offline and come to a decision in a week or two.

Regarding Additional Metadata -

JR: Have received on data set with good metadata - Oregon crest to coast – 20 stations – provides good provenance about instrumentation, calibration, station locations/moves.

Submission agreement – to provide data yearly. Provided by Ron Washman from EPA – sent e-mail to Jared and Valerie.

JL: If others find new metadata sources please pass along to Jared

CM: No additional metadata at the CRU.

Regarding other sources -

ACTION: Jay and Jared to follow up on 3 sources passed along by David Lister.

Also JL talked to Matilde about a possible routine update of the 35 stations from Argentina that we hope to add to GHCN-D this year. She will follow contact the people at INTA (National Institute of Agriculture) and let Jay know.

ACTION: JL to follow up with Matilde regarding the 35 stations and routine updates.

ACTION: Final comments on Annual report from everyone by end of week –

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

PT: Data rescue call held yesterday – low volume turnout. Last time had lots of participation. But this time was in evening and European members difficulty attending.

A number of ongoing activities – mainly European – promised about 100 records – available for v1.1 update and 1 year hence.

PT: Overall initiative – several presentations given by peter and others. Had the SAMSI workshop in Boulder – primarily statisticians and climatologists – to look at bias corrections in new and distinct ways. Hope to finish meeting report within a couple of weeks.

PT: Benchmarking WG- progress cont. algorithm produce 20,000 records with intra and inter-station statistics to create clean worlds.

And Kate remains on target for release of benchmarks in 2015 early. Built off the clean worlds – 10 to 12 worlds released. Most won't have the clean version associated – users only will be able to see the dirty data – a few will be open so users can play with data when they know the answer – all built with different error characteristics – based on random number generation.

Should look like the databank so if can read that should read be able to easily handle the benchmarks.

CM: Will they all be global?

PT: Yes they will be global – came up in discussions in Prague last week – some people concerned they would need global algorithms – No. They can do a subset of 10 or so stations from subdomain of globe. And Benchmark WG would still analyze the results.

CM: any plans to standardize some smaller regions?

PT: not clear – kicked the idea around but if someone is set on doing analysis of California record for instance but we haven't done a case study of CA then they will walk away. Would be self-selecting – don't want to a-priori select regions.

Concepts paper for benchmarking came out – journal – linked off initiatives blog.

## **V. Daily Data, Support to the Monthly Databank and the need for a new umbrella to support multi-element data sets (Menne, All)**

MM: May just use GHCN-D format (stage 1 and 2).

This year will add sources having multi-element holdings keeping in mind the broader vision of a multi-element data set.

Waiting on what others might think about building a multi-element dataset.

VV: German wx serv – precip too (in addition to other groups working on snow, pressure, etc.). Other groups would need to feel ownership.

MM: one vision is for there to be one group where reps from all these efforts (snow, precip, pressure, isti) work together on a true multi-element databank.

VV: at least for precip – much of data closed – can only get maps.

PT: underlying ethos of build design – gpcc and expert team on climate change detect/ind – prepared to accept lack of traceability for better coverage. ISPD and snow cover are doing alternative – sacrificing coverage for better provenance.

MM: the more we build the database the more incentive there is to share.

VV: also makes sharing more easy – don't need to submit to multiple places.

PT: need a land ICOADS – multi-element and no argument of who owns it. Those building GPCC may not relinquish that role easily.

VV: philosophies so different they may not mind – since so much data not open.

MM: those projects could continue – we would try to get data dropped into a central databank – not committing to developing products off of that. Won't have other elements of ISTI such as benchmarking, data rescue, etc..

MM: any suggestions for potential next steps?

PT: obvious thing – if primarily driven by GHCN-D and ISPD in house at NCDC – needs a 2-day discussion meeting where you invite people from these different aspects – could be under CCI under new OPACs– sit people down and discuss things together.

Action: Matt to follow up and determine next step for this. Will check into possibility for such a meeting.

VV: would benchmarking make it attractive for others to collaborate.

PT: one of things about meeting –best practices and opportunities – strengths and spreading best practices.

## **VI. Other Business**

None

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

ACTION: JR – check on number of sources being updated monthly.

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

ACTION: 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.

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

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

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

Action: 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.

Next Calls: All-hands call led by Peter in January. Next Databank WG call in March.