

**Databank Working Group Conference Call**  
**Tuesday, November 26, 2013**  
**1300 – 1400 UTC**

Present:

Jay Lawrimore (JL), Matthew Menne (MM), Byron Gleason (BG), Jared Rennie (JR), Steve Worley (SW), John Christy (JC), David Lister (DL), Colin Morice (CM), Peter Thorne (PT)

Regrets:

Albert Mhand, Vyacheslav Razuvaev, Kenji Kamiguchi, Matilde Rusticucci, Madeleine Renom, Waldenio Gambi Almeida, Meghan Flannery, Albert Klein-Tank, Jeremy Tandy

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

ACTION: ISTI IP comments to Peter / Jay. (COMPLETED)

ACTION: Working Group members to submit metadata and/or parallel measurements now so we can see what exists and inform the data formats.

JR: None received so far.

ACTION: JR to follow up with WG members when review comments are returned. (COMPLETED)

ACTION: JR to undertake stage 3 version 1 release. (To be discussed later this call)

ACTION: Waldenio to assess Brazilian data and advise Jay on potential for gaining additional period of record data.

JL: 4 sets of data being worked by Waldenio. They have been prioritized.

- INMET Conventional stations.. recently began to key from old log books. Until now only data from 1961 for 296 stations. (Priority 1)
- INMET Automated station network (~500 since 2007) (Priority 3)
- INPE's PCD automated station data (~500 stations from 1996 to present.. databank already has 3 years of this data.. Can provide earlier data. (Priority 2)
- Brazilian Regional Network data from automated networks. Manned by states and private companies. Very confusing and quality varies. (Priority 4)

ACTION: DL to make inquiries regarding whether anyone can download data from CLARIS now. (COMPLETED)

ACTION: PT to follow up on location and focus of EarthTemp meeting with eye toward gathering additional source data from Africa.

PT: Focus of workshop will be on data sparse areas with a focus on Africa.. Earthtemp.net – see the website –Any suggestions on participants, experts would be well received. Next call this week.

## II. Status of Databank; Journal article and v1 release (Rennie/Gleason)

There has been a hiccup in getting v1 out associated with configuration management and process review. Code review is currently ongoing. Has delayed the ability to push it out. Holding pattern right now.

Byron stated that the code review is first one in some time. Code is mature research basis. First review 60-70 hours manpower on all aspects of databank code, applying modern coding practices doing best it can. Completed last week. Will shortly look at recommendations and implementation. A number of recommendations have been made. Primarily focusing on improving maintainability of the code.

JR – Expect nothing to change with structure of databank merge as a result of these process reviews.

Beta4 is publicly available and version 1 is internal only until cleared. Compared to beta 4 the version 1 release has updated sources and updated blacklist – no other changes. Just under 32K (31999) stations – much better coverage – some plots on website – similar to before.

Paper accepted for publication – back in august for minor edits. Resubmitted at end of Sept. Now accepted and working on getting it ready to go (page proofs). Working group opinion was strongly don't release until v1 out. Ideally synchronize.

**ACTION – JR/BG to complete review and work with others at NCDC to complete actions required for v1 sign-off by Directors Office.**

SW – How long until release?

JL – Expect 1 to 2 months. But constitutes an uncertain estimate.

Discussion regarding whether to hold the paper. The working group agreed that the paper should not be published until v1 is ready for release.

PT – Recommend that members of the WG be given opportunity to release at the same time public/media information on the databank and ISTI effort, possibly through national media.

CM - Met office may be interested in doing a parallel media story

SW – Possibly at NCAR as well.

JR - CICS will also do something.

PT - And NERSC.

JL – Will follow up with Waldenio when time for release comes.

CM – Albert KT may be interested as well.

JC – Is there an overall press release that could be provided?

**ACTION – JL to send NCDC’s web release. PT theirs from NERSC to WG members.**

III. Update on activities of the Int’l Surface Temp Initiative Steering Committee and Benchmarking progress (Thorne, Menne)

Report from PT

Action items on working group from Steering committee – past due on annual review for WG

**ACTION: Jay to complete WG annual summary report (before Christmas).**

**ACTION: WG members to review Terms of Reference – sunset clauses in all ISTI ToRs – reassess the terms and members to add in or remove.**

**ACTION: Jay to send out e-mail to WG inquiring about each person’s ability to continue as a member of the WG. If not available will ask if there are others who could replace with appropriate areas of expertise.**

Next all-hands call (January) to review initiative as a whole, progress with WGs, revise terms of reference for steering committee.

Other databank relevant info: A letter was sent from WMO SG to all PRs specifically on the databank asking for review by national met services of national holdings... informally heard services have received and are reviewing it. .. one is considering opening a large amount of data.

Put in a proposal for a workshop with SAMSI on developing novel methods for homog that builds off of databank and benchmarks. Accepted and will happen over a 10 day period in July 2014. Cosponsored by NCAR. To be held in Boulder. Announcement to go out in coming weeks. Dates solid now. 8-18 July. Midweek to midweek. Work on one weekend day. This means participants only travel over one weekend. Large number of letters of support across climate, stats, applied math. Hope to convert into participation with PhD and early career scientists. Practically based workshop.

Matt Menne report on benchmarkingWG.

Had a workshop at NCDC last July – chaired by Kate Willett.

3 teams.

Creation – provides the simulated data of databank. Resample GCM output that replicates nature and distribution of data in databank. Those doing homog not intimately involved in creation.

Corruption – adding nastiness, breaks, step change, gradual changes.

Validation – assess the skill of homogenization efforts.

Leading up to workshop was survey of people working in homog- to get sense of nature of breaks. That was basis for nature of breaks to be added. Are some details to work out on how breaks to be added. Next call tomorrow.

Will share what the scenarios will look like on their blog.

JL – how many doing homog?

MM – at least 2 groups early on. NCDC will submit more than one. Berkley group.

PT – some of the avenues explored in the SAMSI workshop discussed above may also pull through.

#### IV. Data Rescue activities (Thorne)

Task team chair of data rescue. Quiet for a year. Hard to reengage the group. Call next week. In theory enough to discuss.

Group has had a little success - possibly a little additional data collected by engaging with data rescue community.

Need to consider whether to fold data rescue committee explicitly into WG, fold task team, or some other option? What does WG think?

JL: Is ACRE the primary source of rescued data internationally? And can one or more members from WG join ACRE as a conduit to bringing in rescued data?

PT: Yes. Recommend JR.

SW: Have been involved with ACRE in the past (ocean side) but willing to reengage with Jared regarding land data.

ACRE just had annual meeting last week and looking forward to next year.

**ACTION: JR and SW to consider following up with ACRE regarding membership.**

CM – can discuss with Jared background on ACRE.

Agreed that WG will defer to judgment of the task team on its call as to whether it wishes to continue in operation or not.

#### V. Data collection activities at the Regional level (All)

DL - CLARIS – can help Jared.

Chinese data – 380 series, quality is good. Homogenized data.

JR – have raw and homogenized.

DL – Possibility of obtaining new data from the Pacific. Most of nations, small nations don't want public domain – could receive but not release.

DL (additional info following call) - Have received 5000+ monthly climate series (Precip., Tx and Tn) from Mexico. I have processed these series to our format and am awaiting station details etc. As with the Mexican files that Jared processed, the crude homogeneity test which uses time-series of annual mean temperature shows a lot of problems. There are frequent high and low excursions that seem to come from the addition of sections of data from different locations/climate zones. I could ask them if we could all share the series. They would be a good test for homogenizing software.

**ACTION – DL to follow up regarding the 5000+ climate series from Mexico.**

CM – Colleague at UKMO recently back from Tanzania and may be able to provide data.

**ACTION: CM – Will follow up regarding data from Tanzania.**

JC – question for Jared – paper on Ugandan data – someone who read recent article has provided new data. Question – should send in same way as previous - station and 3 or 4 time series – or send as is.

JR – easier to do in former since have code set up for that format –

**ACTION: JC to follow up with Jared on additional Ugandan data.**

SW – nothing to report.

JC – longitude fixed on NCAR data??

SW – put an errata message on data. Did not change file. Would have to create new dataset otherwise.

PT – Action on Steering committee on amateur obs the met office is doing – WOW – Bureau of Met to launch a similar site – not sure what action came of that. Web forms to upload – LIKE COOP network for UK and Australia but not managed formally.

**ACTION – Jared to follow up on UK and Australia volunteer programs.**

PT – Is NCAR data in databank?

**ACTION – Jared to look into whether databank is using NCAR data for which errata was posted and if necessary amend blacklist accordingly.**

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

PT – from Steering committee – look at metadata and parallel measurements.

On 4-6 months if have released v1 should be well on way to understanding practices of NRT updates and periodic POR updates.

JC - what meant by parallel measurements?

PT – instruments side by side.. measurement basis understanding, instead of statistical basis.

White literature (met archives) where side by side measurements done – but not shared. Could transform our ability to understand measurements.

**ACTION – Other WG members as able to let others know of the WGs interest in obtaining metadata and parallel measurements.**

SW – contacted colorado state clim. They have it and discussed what is ISTI – should try to revive the discussion.

MM – Nolan has promised to send the data.

JC – RCRN network in US future is highly uncertain but could be used for parallel measurements given their high quality and close proximity.

JC - Going to send a letter to powers that be to encourage continuation of network. Even possibility of losing the 17 stations in AL.

**ACTION - CM – To follow up on possibly parallel measurements available there at the met office.**

### VII. Other Business

Jared to present at AGU (poster)

Jared going to AMS – could bring poster – didn't submit anything since gave talk last year on databank.

Peter to present at AMS.. just an oral on ISTI.

JL: Will remain in touch regarding status of version 1 release, paper, and WG terms of reference.

Will talk again in January with the entire ISTI team.

### SUMMARY OF ACTION ITEMS FROM TODAY'S CALL

**ACTION – JR/BG to complete review and work with others at NCDC to complete actions required for v1 sign-off by Directors Office.**

**ACTION – JL to send NCDC's web release. PT theirs from NERSC to WG members.**

**ACTION: Jay to complete WG annual summary report (before Christmas).**

**ACTION: WG members to review Terms of Reference – sunset clauses in all ISTI ToRs – reassess the terms and members to add in or remove.**

**ACTION: Jay to send out e-mail to WG inquiring about each person's ability to continue as a member of the WG. If not available will ask if there are others who could replace with appropriate areas of expertise.**

**ACTION: JR and SW to consider following up with ACRE regarding membership.**

**ACTION – DL to follow up regarding the 5000+ climate series from Mexico.**

**ACTION: CM – Will follow up regarding data from Tanzania.**

**ACTION: JC to follow up with Jared on additional Ugandan data.**

**ACTION – Jared to follow up on UK and Australia volunteer programs.**

**ACTION – Jared to look into whether databank is using NCAR data for which errata was posted and if necessary amend blacklist accordingly.**

**ACTION – Other WG members as able to let others know of the WGs interest in obtaining metadata and parallel measurements.**

**ACTION - CM – To follow up on possibly parallel measurements available there at the met office.**