

Databank Working Group Conference Call
Thursday, June 5, 2014
1200 – 1300 UTC

Agenda

- I. Welcome and review action items from recent ISTI conference calls (Lawrimore)
- II. Status of Databank and timeline for v1 release (Rennie)
- III. Update on activities of the Int'l Surface Temp Initiative Steering Committee and Benchmarking progress (Thorne, Menne)
- IV. Charting a path for future databank development (All)
- V. Other Business
- VI. 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), Madeleine Renom (MR)

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

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

Done. ORR held last Friday.

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

Sent. Web release now going through a revision and will be online 6/30/14.

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

Completed

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

Sent out on 12/16

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.

Sent out on 12/16

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

Nothing to report.

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

Quality of them bad – on hold. Waiting on metadata. Not looking good. Never got metadata. Realized lots of problems.

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

No progress. Data from a colleague who attended a meeting. Will look into that further.

ACTION: JC to follow up with Jared on additional Ugandan data. (Completed, later 11/26).

John Christy sent Ugandan data on 11/26/13. Is in Stage 1, 2 and latest merge. Monthly temp vales. Took from images of forms – that also have daily data. Will they be digitized? From british website. Will find website.

Pt – possibly under british antarctic survey

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

Peter had reported during November 2013 conference call an 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.

Still need to follow up.

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

Checked that. Was one station with an errata. Our stage2 does not have that. So no need to blacklist.

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

No new data submitted in past 6 months.

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

Parallel measurements over a few months. Elements other than temperature possibly as well.

New actions related to data

ACTION: JC to send web address of Ugandan data with images of daily observations.

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

II. Status of Databank and timeline for v1 release (Rennie)

JR. Past 6 months spent time jumping over hurdles to release databank. Plan was originally October.

Science of alg the same. Did add an ID test in past 6 months. Comparison of stations with good metadata and an ID match – go ahead and merge.

ORR completed last Friday. 1 ½ hour presentation to director and deputy director.

Approval for release on Monday, June 30. Plan to have code along with data online.

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

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

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

Incorporating new sources over time – if any public responses to changes in code. Do remerge every year (annual basis).

PT: How many stations. JR – 32,100 stations. It will change a little with final merge this month.

PT: How many sources. JR - 49 sources in recommended merge (for v1). Frozen about a year ago. Another ~10 sources that will go into next merge - For v1.1 in ~6 months.

PT – when to put out need to say something about ability to contribute new sources and how people can contribute. “we will add new data sources and how to go about contributing”.

PT – some responses to letter under cover to WMO – Swiss data. Later this month expect an announcement from the largest European country.

PT – also doing an announcement on the website. All members need to be encouraged to do the same.

MR – have talked with people from other weather services – person from Ecuador – regional climate service. Working with him to capture all the data can get. Working with matilde and waldenio.

PT – time of release??

JR –10AM EDT.

PT – need to encourage other members to put something out on their website.

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

Benchmarking flurry activity early in year. Then chair had surgery. Only now back to work. Progress slow recently.

Most time of past few months – organize work shop for boulder in july – statistical methods for homoge of sfc temp. proposal of last autumn through SAMSI. 42 participants between stats and climate sc. Largely work based. 10 days in july.

PT – Planning to submit proposal to a Norway funding council with aim to obtain funds for an ISTI in-person meeting.

IV. Charting a path for future databank development (All)

JL – opened up to WG for thoughts from anyone.

Matt presented his thoughts on daily databank. Referencing e-mail and attachments send to WG yesterday.

MM – thoughts included in pdf. Not much benefit to reengineer a new merging algorithm.

Completely leverage off GHCN-D as foundation for daily databank.

About 30 input sources in GHCN-D now.

1) Create Stage 1 and Stage 2 versions of GHCN-Daily sources for ISTI (Note this already happens internal to NCDC as part of the GHCN-Daily processing system—as shown in flow charts)

– Some of these are redundant to the current databank, but most are not

2) The existing NCDC daily merge algorithm will serve to create Stage 3 data (Note this also already happens internally to NCDC, so the only change would be to provide the non-QC'd version of the databank and provide access to code).

3) GHCN-Daily QC algorithms produce NCDC's Stage 4 product (this also already happens).

ISTI side ends at Stage 3. NCDC would pick up at stage 4 (QC'd).

Justification

GHCN-Daily merge algorithm is already analogous to monthly merge algorithm (i.e., uses station metadata matching/data matching)

- The GHCN-Daily processing system is already mature and has a more advanced update/reprocessing cycle than the current monthly databank. The system is also the nexus for official U.S. data processing so it is not likely to be completely de-prioritized by NCDC

- GHCN-Daily has undergone additional scrutiny in the evaluation of the monthly databank without any red flags (e.g., no blacklisting required --similar scrutiny now in process for the precipitation data with comparable results).

- GHCN-Daily is multi-element and these are used jointly in data matching. Would be a step backwards to focus on temperature only.

- BUT, providing the Stage 1 and Stage 2 data permits an alternative merge algorithm if another center wants to commit resources to that.

JC – monthly value archived and also one calculated from daily.

MM – have not done systematic evaluation of that. Likely some differences. We don't know how monthlys were calculated in many cases.

Would be hard to justify NCDRC coming up with a new algorithm. Doesn't preclude another center from doing that if they wanted.

A Few Considerations

Need to overcome some Stage 1 and Stage 2 data formatting issues (e.g., include other elements?)

- Can still work to incorporate sources not yet integrated into GHCN-Daily in the meantime
- Periodic “drop-box” sources are okay, but there are still unexploited web-service databases out there that could be exploited using the GHCN-Daily update/reprocessing paradigm

MM – start this 6 months to a year from now.

Flow Diagrams

MM – operates as drop boxes

MM – if there is a database where there are web services we would like to capitalize on that.

PT – how many daily sources sent to Jared are not in GHCN-D

MM – unsure. Haven't done inventory. Can focus on GHCN-D to bring in sources that haven't yet been brought in.

PT – if attacking in 1 year, in short term prioritize daily sources already there – get them in GHCN-D. so next monthly merge would increase total number via GHCN-D.

MM – yes. Would work with Jared more closely on that aspect.

JC – how to find obs time of max/min. documented? Uncertain?

MM – is treated as uncertain in QC. But through contacts find out country's protocol.. e.g., in Canada (UTC hours that changed in 1960) different than Australia (9AM). is very tricky to rely on Ob time, but there are standards applied country by country.

JL – Big issue brought up by Matt – multi-element. The big 5 (Tmax, Tmin, snowfall, snow depth, precip)?

MM – at least. More than 5 in other countries. Definitely also add Tavg.

Effort in South Am. are exciting because similar to KNMI – regional collections that ISTI can tap into. Experts build regional database. Particularly when updated regularly. ECA&D updated in near real-time for many stations – when combined with other sources get pretty good coverage.

JC – what have you seen over time regarding # of stations becoming available past 10 years.

MM – has not gone down. More and more data being freed up. Benefit of databank using GHCN-D is that won't have big decline in recent data. But we are still underexploiting digital databases. Potential for even more coverage.

PT – consistent with steer from WCRP and global framework for climate services. Would like a single multi-element data holding. Effectively an ICOADS for land. Don't want international pressure, separate from temperature, precip, etc.

So would support efforts to become elemental.

JL – focus for coming year?

MM – as Peter suggested will begin to map out what sources not in GHCN-D. Start to incorporate those. Need to carefully look at that. Would be some temp only stations in GHCN-D but not a problem.

V. Other Business

Peter to attend EarthTemp meeting with a focus on Africa – presenting poster for Jared with the hope of getting new historical data from Africa (meeting to be held in Germany). Will be a number of participants from Africa. Some self supported and others supported by UK.

JC – likely lots of data buried in African met offices.

PT – IEDRO starting to scan microfiche in ACMAD holdings.

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

ACTION: JC to send web address of Ugandan data with images of daily observations.

Following conference call John sent the following.

Monthly forms of daily data from the BADC (British Atmospheric Data Centre) website are images that I believe have, in general, not been keyed in for daily data.

Many images are photographs of the forms that reside in books and include the daily met observations (see attached for Namirembe, Uganda, April 1897).

I looked at every one of the Uganda images (thousands) for my Uganda paper, but only keyed in monthly averages. Some images are extremely clear, some are blank, some are out of focus, some have terrible penmanship, some are faint ... the usual.

The main problem remains for Uganda - virtually no recorded observations between 1979 and 1993. I'm sure there are data records somewhere but have never tracked them down.

Here are Uganda images:

<http://badc.nerc.ac.uk/browse/badc/corral/images/metobs/africa/Uganda>

Africa as a whole:

<http://badc.nerc.ac.uk/browse/badc/corral/images/metobs/africa/>

World:

<http://badc.nerc.ac.uk/browse/badc/corral/images/metobs/>

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

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

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

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.

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

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

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