

20 October 2011 Databank WG Conference Call Summary Notes

1200-1330 UTC

Members In Attendance

Jay Lawrimore (NOAA NCDC, USA, Chair)

John Christy (University of Alabama, Huntsville, USA)

Meaghan Flannery (Australia Bureau of Meteorology)

Waldenio Gambi de Almeida (CPTEC/INPE, Brazil)

Albert Klein-Tank (KNMI, Netherlands)

David Lister (Climatic Research Unit, East Anglia, UK)

Albert Mhanda (ACMAD, Niger)

Jared Rennie (NOAA NCDC, USA) jared.rennie@noaa.gov

Matilde Rusticucci (Univ of Buenos Aires, Argentina)

Steve Worley (National Center for Atmospheric Research, USA)

Peter Thorne (CICS-NC, USA, Chair, Surface Temperatures Initiative Steering Committee, ex officio)

Regrets

Barbara Tencer (Univ of Buenos Aires, Argentina) btencer@at.fcen.uba.ar

Matt Menne (NOAA/NCDC, USA)

Madeleine Renom (IFFC, University of the Republic, Montevideo, Uruguay)

Koji Ishihara (Japan Meteorological Agency)

Vyacheslav Razuvaev (Russian Research Institute of Hydrometeorological Information)

Jeremy Tandy (UK Met Office, Exeter, UK)

The WG welcomed new member Albert Mhanda from the African Centre of Meteorological Application for Development (ACMAD), Niger!

Agenda

- I. Review action items from 30 Jun conference call (Lawrimore).
- II. Update from Int'l Surface Temp Initiative Steering Committee (Thorne)
- III. Overview of Databank Wiki and Databank progress (Rennie, All)
- IV. Strawman for Merging process (Rennie, All)
- V. Potential new sources (All)
- VI. Task team updates. Crowd sourcing and Data Provenance
- VII. Summarize Activities and Next Steps for coming months
- VIII. Other issues

I. Action Items From 30 June 2011 conference call.

- JL to complete poster and begin working on databank pamphlet.
Poster was completed and will be presented at WCRP. There will be three more.
Pamphlet is still pending.

- JL to follow up with Vlad regarding additional progress on mirror databank ftp site at WDC-B.
Databank mirror site is functioning. Daily data are now mirrored as well as the Monthly side of the databank. Mirror site location is <ftp://ftp.meteo.ru/pub/data/globaldatabank/> .
- WG Members to continue pursuing additional source datasets within their region.
To be discussed in Agenda item 5.
- JL / PT Complete and publish the letter and guidance with WG sign off by months end.
Letter and guidance were completed and posted to the Surface Temperature Initiative website.
- JL and JR to work with AKT to look at establishing a periodic update process for the European collection as a test case.
Conference call was held and there are now monthly updates of the European collection.
Script was run this morning.
- NCDC to take action on ISPD lists of data providers and ascertain how to pursue further.
ISPD contacts have led to some additional data. Channel islands and a little more.
- JR to begin calculating daily mean temperature from sub-daily reports in ISPD source data.
Daily means have started being calculated from sub-dailies in ISPD. Flagged with information.
- JL to finalize terms of reference and publish.
Terms of reference have been finalized and published on the Surface Temperature Initiative website.
- PT to soon publish Implementation Plan.
Implementation Plan completed and published.
- Poster will remain a standing item to be used in relevant meetings – ACRE meeting (September) is first case.
Poster was displayed at the ACRE meeting by Kate Willett.
- ALL to try to post the blog.
Blog use minimal.

II. Update on Steering Committee activities.

PT provided an update on steering committee activities.
<http://www.surface temperatures.org/steering-committee> .

III. Overview of Databank Wiki and Databank progress (Rennie, All).

JR provided an overview of the new Databank wiki and progress on Databank development.

Everyone has access to a wiki that describes progress that has been made.
http://editthis.info/intl_surface_temp_initiative/Main_Page . Everyone has read access to this at the moment.

JR described the wiki and how each person can get write access.

Everyone has access to a wiki that describes progress that has been made.
http://editthis.info/intl_surface_temp_initiative/Main_Page . Everyone has read access to this at the moment. This is a free wiki with universal access. Anyone can view, but to edit you will need an account. Request an account from admin which is Jared. Will add some accounts for WG members. Main page is just basic information and links to necessary metadata. Embedded pages list sources and their locations as well as future plans.

Raw GHCND is stage 3, stage 4 is QCed in NCDC's native format.
Daily begins around 1950.
Monthly goes back much further.

Need still to reinject the GHCND sources as stage 1 and stage 2.

Converting daily to monthly using WMO daily to monthly guidance. 5 missing at most, at most 3 consecutive days.

SW: Is there a place that lists the hourly sources that have been uploaded?

JR: We don't as yet have.

PT: We should just take sub-daily and store it for now.

PT: Maps of decadal station coverage?

JR: Yes, can add and make an images folder.

JC: East African stations – most are missing data indicators. Can we nuance so we don't have missing counted?

JR: Yes.

JL: Working on GHCND to monthly?

JR: Already done, should be a stage 2 source.

JL: GHCND original sources are retained, but less so for GHCN-M. Goal is to go back as far as possible and have original format data plus hard copy / image.

IV. Strawman for Merging Stage 2 sources into Stage 3 dataset (Rennie, All).

JL: We are moving closer to the point where we will want to create a merged version of the various Stage 2 sources. Outlining and implementing the process for doing that is not trivial as everyone knows. Some initial thoughts are available at http://editthis.info/intl_surface_temp_initiative/Merge_Program .

We are now grappling with how we should merge the sources. Jared has created guidance on merging data.

JR provided overview of possible merging criteria as outlined on the wiki. Need some criteria, ideas are outlined on the wiki. We have a lot of records 133K+, but these are non-unique. Need to first create a source hierarchy. Max and Min better than avg? Station inventory? Data length? Data back to stage 0?

Merging is based on metadata and data metrics. See the wiki for ideas.

JC: Is there going to be anything done to create a unique set of identifiers?

JR: Need to have a station identifier.

JL: Will be similar to what has been done for daily and monthly GHCN. Identifier will inform on country, WMO id etc. Unique identifier that would index back to earlier parts.

SW: Q. This system will be fully automatic and repeatable?

JR: Yes

SW: This is important so that it can be redone.

SW: Ranking of the datasets might be time variant based upon strengths / weaknesses?

JR: Other way is to create an ensemble with different rankings.

JL: What is the marine side practice?

SW: Unique identifiers and pre-conditioning rules and then a source ranking and record weighting. Grey zone scoring between duplicates and near-duplicates. Fuzzy process. Useful to be reproducible and automated.

JL: Can we see some of this process?

SW: Various whitepapers could be shared with NCDC.

ACTION: SW to share with NCDC.

PT: Please keep an eye on the wiki and provide feedback.

SW: Can we get notification of page changes.

JR: Will investigate and advise.

AKT: Do you anticipate only one merged product? Or more?

JL: Had discussions internally. There will clearly be a need for merging in more than one way.

But should we have one designated?

PT: Yes, this then avoids issues over whether dataset differences arise from raw data or method.

V. Potential new sources (All).

Each WG member provided an update on efforts to acquire new sources of data for the Databank.

MF: Have had good discussions with Janita Pahalad who heads the Pacific Island Climate Prediction Project PI-CPP. Are working to establish with Vanuatu an MOU detailing project goals and data use. Vanuatu has requested assistance with project management and planning. Vanuatu is taking a leadership role within the PI region. We plan on elevating recognition of their spearheading role by creating tangible 'appreciation' of their initiative which is highly visible among other PI countries in the region.

Vanuatu data may be released. MoU dependent. If it requires ministerial engagement then it will take time. The letter and submission guidance were very helpful. The provider had concerns about raw data being used instead of data they quality controlled. Provenance and transparency argument was made and the provider understands.

A certificate of recognition / value to data release would be helpful in conveying to the Pacific Island nations the appreciation that is felt for their contributions of data. Is this possible? WMO contribution to this even better.

Digitization support is required in the Pacific. There is some funding available at BOM for digitization. Exploring establishing MOUs using AUSAID. This is a long-term effort.

Also working on transferring the digitized historical data in ADAM holding tables to the databank. This requires IT involvement on which currently there is a freeze. Will continue to keep the WG updated on this progress and file inventories.

JC: Will scanned images be available for stage 0?

MF: yes

PT: Question about the freedom to use data available on the NIWA site. Can we ascertain what the unrestricted data is and update the databank?

MF: Data from Australia and New Zealand can be included. Data from other Pacific Islands more restricted. Will check and update.

ACTION: MF to chase up with NIWA.

JL: Digitization: How many islands and what period of record?

MF: Pending resolution of MoU before specifics become clear.

JC: Will they digitize everything or just temperatures?

MF: Always digitize everything as a default.

JL: We can certainly do certificates of appreciation. Do they need to be framed?

MF: This would be great. Framing can be accomplished by BOM.

ACTION: JL will take forwards.

PT: We can see at WCRP OSC whether we can get CCI, GCOS and WCRP co-signatories.

AKT: With WMO Congress coming up next year, can we make this more substantial and include some materials?

PT: Number of potential pathways.

David Lister provided update from the UK.

DL: Not particularly successful. Bangladesh was unsuccessful – regarded as a state secret. There is some Ghanaian data in hard copy some of it disintegrating.

PT: Where is it?

DL: in Phil Jones Office.

PT: Suggest that you contact IEDRO.

ACTION: DL to chase up with IEDRO.

JC: What do you do if you find in your possession a hard copy of the data that is otherwise proprietary? Is it still private?

JL: They must have been okay with releasing in the past.

SW: There should be a copyright statement in the publication.

JL: Still looking at crowdsourcing possibilities for converting the millions of imaged forms to digital data. There is strong support from NCDC director's office for this. Expect effort to begin sometime in FY12.

PT: Can we go back with some of the Brazilian stations Waldenio provided?

Waldenio: Present issue is resource. Are doing what we can to improve those holdings.

CPTEC have only digital. But Met service is digitizing. Data policy is an issue. Optimistic about improving.

JL: Would a letter of request help?

Waldenio: Yes, probably.

ACTION: JL to follow up with Waldenio on letter.

JC provided an update on Ugandan data that is going to be digitized. Wait until whole thing done? Or just send what you have?

JL: Send when you have. Processing overhead is minimal.

ACTION: JC to send data.

SW provided an update on data source possibilities at NCAR. Library at NCAR is looking at holdings and has contract to get some imaged and OCR'd. Looking at examples from Africa and Antarctica. Big Q is how much is in there that is new. A few new data. Greenland. Will continue to pursue this and will report progress on next call.

AKT provided an update on possible new data sources associated with ETCCDI meeting to be held in December. This is being led by ACMAD. AKT will raise the issue and inquire about possible participation and submissions to the Databank. MR will also be present.

VI. Task team updates.

Crowd sourcing. <http://www.surfacetemperatures.org/databank/data-rescue-task-team>.

PT provided an overview of the crowdsourcing task team recent activities. Continues to run into problems associated with the need for funding for imaging and digitization of the many sources available.

JL: We keep getting mixed messaging on climate data modernization program from FY13 onwards. There may be new opportunities in FY13 but will not know for some time.

Data Provenance and version control. There is one new document on the pros and cons of the use of UIDs at <http://www.surfacetemperatures.org/databank/provenance-and-version-control-task->

[team](#) . There have been discussions on UIDs similar to ICOADS chosen route. Still unresolved. Summary available online.

JL: There remain questions regarding whether there really is a need for that level of granularity of identifier flags for land surface temperature records.

SW: Agree that while the unique nature of marine observations, the UIDs may not be helpful to land surface observations. The problem needs more consideration. We should look at more and talk again. Will discuss further.

VII. Summarize Activities and Next Steps for coming months

JL: Goal is to do an initial release in April next year, so timelines and goals are driven by this.

PT: Next discussion should be based upon progress to making a stage 3 merge. Will want to have a call to discuss the merging process before it is finalized. Progress expected on this and next call may be possible within two months or less.

ACTION ITEMS

SW to share ICOADS white papers on merging process with NCDC.

JR to investigate possibilities for automatic notification of wiki changes to working group members who are interested.

JR and others to continue working on merging process for Stage 3 Monthly data and to report back to the WG on strawman.

MF to continue working with Pacific Island partners and to report back on progress next call.

MF to follow up and determine what data sources provided by NIWA can be included in the Databank and made freely available.

JL to work with Graphics team on developing a certificate for country recognition.

PT to coordinate at WCRP OCP to determine if co-signatories can be made by CCI, GCOS, and WCRP on the certificates.

JL to prepare letter for high level request for support from Brazil INMET. WGA to follow up with JL on the issue.

DL to follow up with IEDRO to determine if they would be able to assist with imaging of the forms in Phil Jones' office.

DL to follow up with JL and JR on other sources from the UK.

JC to send Uganda digitized data as it becomes available.

SW to continue to pursue data associated with NCAR library digitization efforts.

AKT to discuss Surface Temperature Initiative with participants at the ETCCDI workshop in December and investigate potential new participation and possible data sources.

VIII. Other issues.

None.