

30 June 2011 Databank WG Conference Call Summary Notes

Members In Attendance

Jay Lawrimore (NOAA NCDC, USA, Chair)
John Christy (University of Alabama, Huntsville, USA)
Waldenio Gambi de Almeida (CPTEC/INPE, Brazil)
Albert Klein-Tank (KNMI, Netherlands)
David Lister (Climatic Research Unit, East Anglia, UK)
Matt Menne (NOAA/NCDC, USA)
Madeleine Renom (IFFC, University of the Republic, Montevideo, Uruguay)
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)

Also in Attendance

Jared Rennie (NOAA NCDC, USA) jared.rennie@noaa.gov
Barbara Tencer (Univ of Buenos Aires, Argentina) btencer@at.fcen.uba.ar

Regrets

Meaghan Flannery (Australia Bureau of Meteorology)
Koji Ishihara (Japan Meteorological Agency)
Vyacheslav Razuvaev (Russian Research Institute of Hydrometeorological Information)
Jeremy Tandy (UK Met Office, Exeter, UK)

Agenda

- I. Review action items from 5 April conference call
- II. Databank development progress and questions for WG
- III. Review data submission guidance and cover letter
- IV. Review databank working group Terms of Reference
- V. Review Steering Committee's IP, particularly Databank milestones (pgs 7-8)
- VI. Task team updates
- VII. Summarize Activities and Action Items for coming months
- VIII. Other issues

I. Action Items From 5 April conference call.

1. WG members have a goal of providing at least one new source of data in Stage 1 format (and Stage 2 if practical).
 - Specific commitments
 - a. MF: from Pacific and NZ.
 - b. WA: Brazilian CPTEC data. And will talk to Brazilian Met Service. **(Initiated)**
 - c. MRu: Will look into availability of CLARIS data from River Plate
 - d. MRe: A number of stations from Uruguay, some recently digitized from beginning of 20th century. **(Initiated)**
 - e. MM to provide data for US/N.A.

- f. AKT from European collection. **(Completed)**
- g. SW from NCAR collection.
- h. JC data from East Africa.

JL: Brazil stage 2: In work progress based upon examples of stage 1 to stage 2 and feedback. Helps us out to have both.

- 2. PT to develop data request letter for signature by CCI, GCOS Steering Committee, Databank, and Steering Committee leads. **(Attached)**
- 3. JL to address pamphlet and poster to be produced to be produced by NCDC graphics. WG members to review. **(Not completed)**
- 4. JL to document and provide guidance for data submission **(Attached)**
- 5. JL to develop terms of reference for the Databank. WG members to iterate and modify. **(Attached)**
- 6. JL to refine Data Provenance Tracking flags and iterate with WG. **(Ongoing)**
- 7. JL to consider additional options for databank mirror sites, such as WDC-B. **(Initiated test site w/ WDC-B)**
- 8. PT to follow up with UK rep to WMO Congress **(Completed)**

PT: Yes, text on the surface initiative made it into congress output including data submission request and generic email address data.submission@surfacetemperatures.org to submit to.

- 9. PT to establish a blog and e-mail alias for the WG. **(Completed)**

Related Issues

With Slava's help and the efforts of Dr. Vladislav Shaymardanov (Vlad) at RIHMI-WDC we now have a second databank site with the same structure as the one at WDC-A for Daily data. <ftp://ftp.meteo.ru/pub/data/globaldatabank/daily/> . More work is to be done to establish a maintenance schedule to keep it updated on a routine basis and to add the Monthly data structure.

World Weather Records: Request that this process be transitioned from Decadal to Annual updates. NCDC produced a pamphlet to support this issue for WMO Congress. Was presented as an intervention during the WMO Congress. And subsequently voted in favor of this. Plan is to complete the 2001-2010 WWR edition and then begin addressing the annual updates.

JC: Is that going to be printed?

JL: No, just electronic aim is to make it more routine.

SW: Are these just delayed mode summaries?

JL: Yes, largely delayed mode monthly effort to garner data, was done decadal since the 1920s.

DL: These are a great resource but quality wasn't perfect – some systematic issues in the 1990s WWR collection.

WCRP Open Science Conference: Our abstract on the Databank was accepted as a Poster. This along with other posters on the International Surface Temperature Initiative will be presented in a cluster. Peter Thorne will also provide an oral presentation.

PT: Will request help nearer the time with databank aspects of the oral presentation.

Article of the International Surface Temperatures Initiative will be published in BAMS as a meeting summary from the Exeter workshop.

JL: Does everyone have a copy?

II. DATABANK DEVELOPMENT PROGRESS

Recent Daily Data Submissions

Japan (Koji Ishihara)	Daily data for 157 stations.
Europe (KNMI; Albert Klein-Tank):	Daily data for 1177 stations.
ISPD (Int'l Sfc Pressure Databank):	Daily data for 13 stations in Tunisia&Morocco.
Pitcairn Island (Peter Fisher)	Daily data for 3 sources of same station

Previous Daily Data Submissions

Spain (Manola Brunet)
US Forts (CDMP)
Vietnam (CDMP)
Mexico (CDMP)

Recent Monthly Data Submissions

East Africa (John Christy)	Monthly data for 112 stations.
Europe (KNMI; Albert Klein-Tank)	Monthly data for 1551 stations.
Japan (Koshi Ishihara)	Monthly data for 157 stations.
Antarctica South Pole (Matthew Lazzara, Space Science and Engineering Center at the University of Wisconsin-Madison)	Monthly data for 1 station.

Previous Monthly Data Submissions

Colonial Era Archive
World Weather Records (Through 2000)
Antarctica (SCAR) (Antarctic Climate Data SCAR READER Project website -- <http://www.antarctica.ac.uk/met/READER/>, downloaded on 9 Feb 2011)

JL: Uruguay and Brazil data submissions in progress.

QUESTIONS FOR DATABANK WG:

1. Calculation of Daily Mean Temperature from Sub-daily temperature observations.

Most of the ISPD source data do not include daily max/min/mean temperature. There is only one so far we've been able to identify with daily max/min/mean temperature. The others contain sub-daily observations; some hourly, some synoptic, some others such as tri-daily.

Question: Does the subgroup recommend we use these sub-daily observations to calculate daily Mean temperature and with what stipulations?

JC: Key is the provenance tracking here. As long as the calculation method is recorded and sufficient number of hours are included in the daily average. Need 4 for Tmean.

JL: Forts data is similar

DL: Not a trivial issue. It is important to maintain consistency of the record – calculating daily means in same manner through time.

JL: But historical data is not consistent and in many cases we do not know how the daily mean or monthly mean temperature were calculated.

DL: Australian example gives a difference between the average of hourlies and average of max+min/2. At one point CLIMAT data were transmitted as max+min/2 and then switched to average of hourlies.

JL: Agree. But with databank we want to leave decisions to some extent to users in developing the stage 3 merged dataset. If provenance is provided on all values the various users can make decisions regarding which data values to include.

SW: Yes, track the provenance. Think about user interface GUI to aid users.

JL: Will need to add additional provenance tracking flags for calculation of daily means from ISPD source.

The number of hourly temperature values vary by source, and sometimes even vary by station.

We have attempted to create daily averages and convert those sources to stage 2. Only creating an average if there was sufficient data, which depended on the data set.

For example, if temperature was recorded on an hourly basis, we required at least 20 valid values before an average was calculated. If there were only 4 recorded values per day, we required all 4 values be non-missing before averaging.

Are these criteria too stringent?

If we use sub-daily data to get daily mean temperature, we will need to add another data provenance flag. Perhaps something like a "calculated flag" where:

100 = data values are original

101 = data values were calculated from sub daily data (at least 3 obs available)

102 = data values were calculated from sub daily data (at least 10 obs available)

ACTION: JR to begin calculating daily mean temperature from sub-daily reports in ISPD source data.

2. There is currently no mechanism for making periodic updates to the data in the databank. Do we need to be concerned about this at this point? Should we be putting some effort into establishing update processes or continue to focus solely on obtaining new historical data sources?

JL: Assume that many of the sources are updated

AKT: Not just additional points but additional stations.

JL: GHCND has various methods of updating source data.

MM: Daily weekly and monthly updates. We are in a position to automatically update on whatever appropriate scales.

JC: How often is the data re-corrected at source?

MM: Tiny %age of total but can be a large number.

JC: Concern that numbers change.

PT: Use of UUID for each data value is essential to complete data provenance.

MM: We version the entire GHCN-D dataset by date but not each data value.

SW: Not quite at UUID yet for ICOADS. Currently only release by release number no update between release. Need to put the UUID on stage 2 data at a minimum.

JC: How do you associate changes?

PT: You could have an associated metadata file in theory.

SW: Will share the ICOADS process development documentation.

ACTION: JL and JR to work with AKT to look at establishing a periodic update process for the European collection.

3. We have largely obtained the low hanging fruit – from this point forward more effort will be required in obtaining additional sources of data. Are we taking the right approach by having the working group members reach out within their regions? Are there other possibilities we should consider?

PT: ISPD data providers largely submitted pressure data but very likely also have temperature data that could be provided.

JL: We will pursue obtaining a list of the providers as a first step toward making contact for temperature data.

Action: NCDC to take action on ISPD lists of data providers and ascertain how to pursue further.

4. What are the general opinions regarding the progress the Databank has made to date?

SW: Pretty good progress for unfunded mandate. Need to recognize this.

III. DATABANK SUBMISSION GUIDANCE and COVER LETTER

We will review the attached submission guidance for additional comment and suggested edits.

JC: What about the submission of the image file or the hardcopy?

JL: We perhaps need to include this as a desirable. Will nuance the text in this regard.

JL: Should we make metadata required? Matilde reports that Uruguay has non-standard reporting practices for example so we need to capture these.

MRu: Yes, Uruguay precipitation and temperature is effectively separated by a day due to this practice and that is important to know. Max/Min temperature are reported on the day of the observation, but precipitation observation taken at 0700 is reported as the previous days observation.

JL: Surely not just Uruguay. So do we want to require the additional metadata information or add some station information?

AKT: Specify requirement but allow user to select data based upon provenance.

MM: Don't want people to avoid submitting because of overhead or because they don't own the data.

DL: Could be helpful to request provider include their national convention for observation recording in the required metadata file.

PT: Also could change "optional" to something stronger such as encouraged or strongly encouraged.

SW: Agree and add reasoning to this document to justify that effort on their part.

JL: Providers can periodically update data through the ftp site.

ACTION: JL to modify data submission guidance letter and return to WG for final comment and then circulate final for use by regional members in acquiring additional data.

Likewise with the cover letter.

PT: IEDRO DVD link is waiting on their input on technical host.

ACTION: PT / JL to approach suggested co-signatories

IV. DATABANK WG TERMS OF REFERENCE

We will review the attached Terms of Reference for additional comment and suggested edits.

JC: If someone is running code and finding useful information. How is this being communicated back

PT: Not clear but clearly this is important and will be necessary to associate back to the raw data.

SW: Should we be thinking about expanding the network?

JL: Need a member from RA1. Do we need to add more members beyond? Koshi is looking to find a replacement following a job shuffle. China member?

ACTION: PT to follow up on Chinese representative.

V. STEERING COMMITTEE IMPLEMENTATION PLAN

JL Provided brief summary of databank portion of plan and reviewed the Databank milestones on pages 7-8.

VI. TASK TEAM REPORTS

Data rescue and Crowd sourcing (Thorne)

Mainly coordination thus far. List of known unknowns and others is hosted on BADC (through ACRE). Need to think as to how to pull through. May be crowdsourcing funds available.

JC / JL crowdsourcing would be useful.

JC: Is it just set up?

PT: No, also some in-house support for science pull through and user engagement.

Data provenance and version control (Lawrimore)

Little update since last call. UID is clearly key.

Action: JL and JR to begin investigating use of UID with data provenance team.

VII. ACTIVITIES AND ACTION ITEMS FOR COMING MONTHS

ACTION ITEMS

- **JL to complete poster and begin working on databank pamphlet.**
- **JL to follow up with Vlad regarding additional progress on mirror databank ftp site at WDC-B.**
- **WG Members to continue pursuing additional source datasets within their region.**
- **JL / PT Complete and publish the letter and guidance with WG sign off by months end**
- **JL and JR to work with AKT to look at establishing a periodic update process for the European collection as a test case.**
- **NCDC to take action on ISPD lists of data providers and ascertain how to pursue further.**
- **JR to begin calculating daily mean temperature from sub-daily reports in ISPD source data.**
- **JL to finalize terms of reference and publish.**
- **PT to soon publish Implementation Plan.**
- **Poster will remain a standing item to be used in relevant meetings – ACRE meeting (September) is first case.**
- **ALL to try to post the blog**

VIII. OTHER ITEMS AND NEXT CONFERENCE CALL

Next call in late Sept / early Oct.

JL to send out a doodle request in early September.