

International Surface Temperature Initiative call June 4th 2013

Open call for all participants in ISTI committees, working groups and task teams

12Z (8 EDT, 13 BST, 14 CEST etc) June 4th 2013

Present on call / etherpad: Peter Thorne (PT), Victor Venema (VV), Kate Willett (KW), Andrea Merlone (AM), Jay Lawrimore (JL), Claude Williams(CW), Blair Trewin (BT), Jared Rennie (JR), Manola Brunet (MB), Matt Menne, Rob Allan, Colin Morice, Robert Dunn, Akiyo Yatagai, Antonio Possolo, Ian Jolliffe, David Lister, Waldenio, Stefan Bronnimann, Albert Mhanda

Apologies in advance: Xiaolan Wang, Albert Klein Tank, Lisa Alexander, Richard Chandler, Juerg Luterbacher, Rod Hutchinson, Matilde Rusticucci, Madeline Renom

1. Call background

This call was intended to engender broad strategic discussion amongst all ISTI participants and act as a means to update all participants as to where we stand in various aspects of our work and enable constructive criticism and ideas.

Resulting major action items arising:

ACTION: Everyone to provide track change comments on the IP document draft to PT by 14th June. Refresh will also inter-alia take into account the outcomes of this call and any additional suggestions received in the interim.

Resulting minor action items arising:

ACTION: PT to chase GCOS secretariat for an update on the letter requested to be sent to PRs and GCOS focal points at AOPC

ACTION: PT and AM to add meteomet text to IP refresh with some output that is ISTI relevant.

ACTION: KW to contact Manola and Andrea about reporting to the Benchmark WG on this Type B uncertainty and intercomparison knowledge during a weekly teleconference.

ACTION: Databank Working Group to add some realistic items into IP on metadata rescue and preparation.

ACTION: PT to follow up with GCOS secretariat regarding request to garner metadata for GSN and RBCN network stations and include JL.

ACTION: SB to ensure ERA-CLIM phase 1 data are submitted to databank.

ACTION: Steering committee, databank working group and data rescue task team to organize calls before end of July.

2. Status updates (since last all members call last boreal(!) winter)

2.1 Steering committee - Peter Thorne

The steering committee has continued to meet occasionally. The annual report was completed and published. There has been some progress in publicising the initiative and some oversight given to the working group activities. Steering committee members have constructed and shared a spreadsheet of potential investigators who may be interested in participating in analysis of the databank and the benchmarking effort but this is stayed pending formal first version databank release.

PT attended the GCOS Atmospheric Observations for Climate. Relevant minutes taken from the published report are:

AOPC was pleased to see arrangements for reporting of daily climate observations moving forward quickly, following the recommendation made to CBS through its Implementation/Coordination Team on Integrated Observing Systems (ICT-IOS) agreed by the Panel during its 2012 session. The Panel was grateful for the efforts of NCDC in working with partners to develop this via BUFR-formatted messages. It encouraged approval by CCI, recognizing that tools and support for implementation would be needed. GCOS should promote implementation by the operators of GSN stations in particular.

AOPC thanked Dr Thorne for his presentation. The Panel was very interested to see the progress that had been made. It noted the substantial increase in the amount of data that has been made available for analysis, and appreciated being shown early results from such analysis. Although this indicated little change to estimates of global trends, significant local changes were emerging.

AOPC also welcomed the new project to build a database of the parallel measurements needed to study daily inhomogeneities under the leadership of Dr. Victor Venema.

Noting that support had been given by CCI and WCRP, as well as GCOS when the ISTI had been launched, and that it was desirable for there to be renewed national contacts for local verification of the merged database and to find new data sources, AOPC requested the GCOS Secretariat, in liaison with its CCI and WCRP counterparts and with Dr Thorne, prepare a letter to Permanent Representatives (PRs), with copy to be sent to national focal points. In addition to requesting feedback on the merged database and any additional observational data that could be found, the letter should explain how the initiative was now beginning to make data and information available to the countries in return. It should also draw attention to the new activity collecting parallel measurements, and call for any national data holdings that would contribute to this project to be made available.

ACTION: PT to chase GCOS secretariat for an update on the letter requested to be sent to PRs and GCOS focal points at AOPC

Publicity for effort:

Talks at North Carolina State University, SAMSI massive datasets workshop

Talk at WMO Climate Watches/Climate Monitoring regional workshop for eastern/southern Africa (South Africa, April 2013)

2 posters at EGU meeting in Vienna

2.2 Databank Working Group - Jay Lawrimore

As reported in the last conference call a beta release of version 1 of the Stage 3 databank has been available since October 1. Feedback and development continued and a third update of the beta release was completed in response. This release remains available at <http://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/> .

Updates made as part of the third update are available at ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/updates_to_beta3.pdf. The changes include additional of a blacklist of stations to correct errors or to withhold stations whose quality or provenance is suspect. Some code changes were made to improve the quality of station merges, and the format of the stage3 data was changed to ensure consistency among the various stages of daily and monthly data. The fourth change made was the addition of metadata output from the merge program to provide users with information regarding each candidate station comparison.

The NCDC in situ configuration control board reviewed progress on the databank on 12 March and recommended full operational release of version 1.0.0 following submission of a peer reviewed article and acceptance with only minor revisions.

Although a brief report on the databank was published in EOS several months ago full details on its development will be included in the future article. Jared Rennie with the assistance of all databank working group members completed this article in April and submitted it to Geoscience Data Journal, a new open access journal. Review comments are pending.

Andrea Merlone: MeteoMet series sent. MeteoMet WG4 is working on the first attempt to introduce type B uncertainties in those data.

PT: Jared Rennie to confirm he has these

JR: Data was placed on our FTP site by Peter Pavlásek, however some FTP issues over the past few weeks led to data deletion. Requested Peter to re-send the data.

PT: Can we clarify for the generalist audience what the type B uncertainties involve? How might they be useful for a. homogenization efforts and b. benchmarking?

AM: Data is from several European countries. Have started to see whether we can append some information based upon the instrumentation and metadata - an instrument based uncertainty estimate. Uncertainty budget is calculated through a series of contributing factors based upon physical understanding. There are a number of resources being given to this problem. Expect to deliver first half of 2014.

ACTION: PT and AM to add meteomet text to IP refresh with some output that is ISTI relevant.

VV: How is this distinct from benchmarking?

AM: This is purely physically based instrumental uncertainty. Not sure that we can be comprehensive, but can give uncertainty expectation based upon measurement technology knowledge. In deep past that knowledge is incomplete.

MB: Working on carrying out field comparisons in long-standing observatory facilities in Spain and N. Italy with parallel measurements between calibrated new instruments and old existing instruments, to assess likely biases and uncertainties. Also includes trying to assess adjustment uncertainty for instrument change from old to new systems.

ACTION: KW to contact Manola and Andrea about reporting to the Benchmark WG on this knowledge during a weekly telecon

PT: Work is nearly finalized to in addition create a netcdf CF compliant version of the individual station files. Main issue outstanding is with the netcdf attribute values for the flags. Although the attributes can have arrays for the values it doesn't seem possible to pass the resulting flag descriptors as an array. Investigations are ongoing and I am working with the CF governance team to ensure we get this right.

PT: Databank currently is Tx (maximum), Tn (minimum) and Tm (mean). Should we also calculate and populate a Tdtr (diurnal range) field for all months with Tx and Tn as the default? Am (optionally) calculating in the netcdf file versions. Issue here is what if Tx arises from say 28 days and Tn from 25 days and (worse) the 25 is not entirely a subset of the 28?

JL: Wait and leave it up to users?

BT: Tm is not a derived product in many cases whereas DTR is.

General agreement not to release a derived Tdtr field at this time.

VV: How much station history metadata do we have? And do we have a definition for a machine readable format for them? Sounds like a complicated, but very important task to make automatic homogenization algorithms more powerful.

BT: very hard to develop an exhaustive 'standard' format for metadata - for example, site coordinates can handle site moves but not changes that occur at a location in situ.

VV: I agree, it will be hard, but it is also important and an international initiative as the ISTI would be ideally suited to try to formulate standards. If only because we also have informatics hacks in our group. We can start modest: list potential dates and periods with any kind of inhomogeneity, without much specification, list outliers. And list dates that seem like outliers or breaks, but are real. It should be some sort of extensible format.

PT: This item could be added to the WG list for the databank if they agree. We'd need to craft words and realistic actions to accompany.

JL: Agree that this could / should be a focus.

ACTION: Databank Working Group to add some realistic items into IP on metadata rescue and preparation.

JL: Who are candidate countries? Can we start with low hanging fruit?

VV: Alpine countries are relatively complete.

PT: From AOPC meeting there was discussion about asking national focal points and PRs for metadata on GSN and RBCN network stations.

ACTION: PT to follow up with GCOS secretariat regarding request to garner metadata for GSN and RBCN network stations and include JL.

CW: Station history information outside of the US is limited. All of the stations have at least lat/lon/station name in the station list. For the stations that are a merge of different sources, there is a list of the dates of the merges and the source. As for formats, NCDC has used everything from the Master Station History Records inside the US (very complicated) to a record for each candidate change date in the form of "Station ID, Date" (pseudo station history for the NCDC Benchmarking study).

PT: Can we build off the (most comprehensive) US metadata as a standard?

CW: Tis a grand and noble cause, but you have to get countries to let go of their data...

PT: Agreed. But if we give them a template to work from they may be more willing than just saying you're on your own. And it makes sense to start from the most comprehensive current set of machine readable holdings rather than have to down the line reverse engineer these to some new standard.

CW: Aye

VV: I have no idea how good the US standard is, but if informatics can provide us with a standard that fits to other more general standards (XML, semantic web) and is extensible (maybe a hierarchy of detail), it might be worth to make a new standard and convert the US metadata in that standard.

CW: Problem is, we need to know what information other offices have in order to start. Does this indicate we need some type of survey to find out the most common information? Our histories are built from an old NWS form (10-A & 10-B) and has information like distance and direction from post office - which was a standard (if no other location data was available) back in the early part of 20th Century.

2.3 Data rescue task team - Peter Thorne

This has been silent for some time. Not as a result of the members but rather that the chair's time has been squeezed in too many ways and something had to give. We should aim to resurrect this task team's activities.

We have an IEDRO Data Rescue Team heading to ACMAD in Niamey, Niger next week to establish the West Africa Climate Data Rescue and Digitization Facility at ACMAD. Four microfiche scanners will provide JPEG images from the nearly 2,000,000 DARE-I microfiche frames over the next several months. Phase II of the program will provide for the digitization of those data some of which go back to the mid 1800s.

Juerg Luterbacher: We will go on our digitisation efforts and I will soon send you a long sub-daily series from Tanarive and other smaller ones

Manola Brunet: In addition to Juerg's comment, under the EURO4M project linked to the WMO/MEDARE about 1,2M daily values over 68 North Africa and Middle East locations have been digitised, quality controlled, merged (with recent available parts to compose longer Tx/Tn series), adjusted and will be made available to the databank once possible (not later than the end of this year, and hopefully earlier). So, even the silent period, things are progressing in the team.

MB: No later than end of this year these daily values will be able to be provided.

JL: Can provide raw as well as QCed and adjusted?

MB: Yes, certainly.

MB: There is follow on funding which will be used for more data rescue.

RA: On the IEDRO effort - the microfiche inventory is complete for French colonial, but British colonial regions not from 1950 onwards. UK have these in archives and have been imaged. Juerg's effort is being prioritized by ACRE to look at digitizing these to complete the set. Bottom line: there are some additional data that could augment EURO4M that have been imaged.

SB: ERA-CLIM phase 2 has been funded. Surface temperature will be digitized when done.

ACTION: SB to ensure ERA-CLIM phase 1 data are submitted to databank.

BT: Important to promote digitizing everything as ensures better understanding of the data and the homogeneity issues.

PT: Long-term aspirational goal is to create an integrated set of holdings (multi-elemental, multi-timescale, integrated) but that is resource limited presently. It is something we should not lose sight of but is a massive undertaking given the fractured nature of the archival landscape.

VV: Prioritise additional elements for which there are no global holdings yet?

2.4 Benchmarking and Assessment working group - Kate Willett

We have established three subgroups to focus on simulation of realistic 'clean' station data, design and implementation of error-models and assessment of homogenisation algorithms against the benchmark station data. (Team Creation, Team Corruption and Team Validation respectively).

Team Creation have been working with vector auto-regressive models to replicate high frequency station characteristics and using GCM gridboxes to provide low frequency characteristics. This work is close to completion.

Team Corruption have been defined systematic biases, seasonal cycles and remaining uncertainty as big questions to focus on with the error models. A set of 'blind' and 'open' worlds are being designed for the version 1 release.

Team Validation are working with assessing both the ability of algorithms to detect when inhomogeneities occur and also their ability to return the data to its 'clean' form from the station to regional to global scale.

Funding has been obtained for a PhD student to work on creating daily benchmarks - we're focusing on monthly here. Funding has also been found to run a 3 day workshop at NCDC which will bring together a small number of the working group for both finalising concepts and physical development of the benchmarks.

Overall our team is behind schedule but progress is being made.

3. Refreshed Implementation Plan

The original Implementation Plan (http://www.surface temperatures.org/steering-committee/Surface_Temperature_Initiative_Implementation_Plan_version1_release.pdf?attredirects=0) was drafted by the steering committee and ended up partially achieved. It was agreed by the steering committee members that a new Implementation Plan should be created to cover 2013-2015.

Items to consider:

1. Membership of groups, role of groups, whether new groups are required current ones should be restructured etc.

This is potentially an opportunity to refresh some of the groups if people feel this is appropriate.

JL: WG on databank is a pretty solid group. May want to add some folks. People are mostly engaged and making significant contributions.

KW: We are refreshing the benchmarking WG. Difficult for people to clear the time, but with clarity of requirements its getting easier. We need more software experts somehow. How do we do this without funding?

JL: Task team on metadata given preceding discussion?

PT: Happy to support formation, leave this up to databank WG to decide

2. Realism of timelines

IJ: On validation part of the IP happy to lead so long as the corruption is done. The issue is with software development and moving it forwards – timeline realism implications if we get no resource to support someone developing the software to implement what the team comes up with.

KW: this is a problem more generally I think as we don't really have any software experts and these are difficult people to get on board. We will cope with what we have but it does take longer. I'll have a think about this.

IJ: I'll also think about members of the verification community who might be approached.

3. Whether there are big ticket items missing as things stand.

CW: Funding for face to face meetings. Priority to try to get funding for a general meeting of all participants or subsets of participants.

ACTION: Everyone to provide track change comments on the IP document draft to PT by 14th June. Refresh will also inter-alia take into account the outcomings of this call and any additional suggestions received in the interim.

4. Other items of interest

Please place any items of general interest here. Include things such as upcoming meetings, suggestions for things we could / should be doing etc. etc. here. Possible items to discuss:

workshops / conferences

Please advise of any relevant workshops/conferences we should be aware of:

AM: how about ISTI to endorse the 2014 workshop on Metrology for Meteorology and Climate?.

PT: Good Point. Please provide some brief details if possible. In theory I think this makes sense but I know more about this than other participants so a little more detail might help here.

AM: Will be in first week in June 2014, 2-3 days in Slovenia. Bringing together people from meteorology and metrology. Some funding exists through EU. This workshop will bring together groups of people.

BIPM co-sponsoring.

ISTI will endorse.

VV: IMSC at the end of this month will have two talks on ISTI by Kate. Kate's PhD student Rachel will talk about creating daily benchmarks. Her work will benefit ISTI in the long run.

VV: It sounds like many ISTI members will be at IMSC. Would it make sense to use this opportunity for discussions?

Kate Willett, Victor Venema, Ian Jolliffe, Rachel Warren...

KW/IJ/VV/RW to meet up while there

KW: Benchmarking workshop 1st-3rd July now confirmed. A small and hands on meet up to get the benchmarks up and running as far as possible in three days.
Daily reporting to occur during workshop so online involvement is possible/encouraged.

Blog and website

Please provide any constructive criticism and / or ideas here.

VV: In the beginning the blog was read a lot and we got a reasonable number of comments. That momentum seems to be gone. A recent discussion initiated by the benchmarking group on the homogenisation email list gave a lot more email response as blog comments. Maybe we should accept that most scientists do not like to discuss in the open. Would it be an idea to have an additional distribution list for everyone in the ISTI and every interested? Maybe also a newsletter would work better at this more quiet stage as the blog/homepage.

VV: Newsletter would be similar to what is currently on blog.

KW: Important to keep the blog and update and engage people.

KW: Twitter account? I've been tweeting already - sorry! hashtag?

VV: #surftemp ? #isti is already used a lot.

KW: Benchmarking group can contribute more to this with regular updates on progress.

JR: Twitter can be very powerful

PT: Need section in the IP on communications.

Engendering new dataset efforts

There is an open access googledocs link that contains a master listing of potential parties to engage in an attempt to engender new dataset efforts. This can be shared upon request.

VV: Main stumbling block is resources. Funding is important. Otherwise mainly build through personal contacts.

PT: ISTI will happily write letters of support for relevant activities.

5. Any Other Business

We should have meetings of databank WG, data rescue and steering committee in next two months.

ACTION: Steering committee, databank working group and data rescue task team to organize calls before end of July.