

ISTI Steering committee call

11/9/11

8am EST, 13Z, 13UK, 14CET ...

Present on call: Peter Thorne(PT), Michael de Podesta (MdP), Andrea Merlone (AM), Chris Merchant (CM), Jay Lawrimore (JL), Blair Trewin (BT), Albert Mhanda (AMh), Richard Chandler(REC)

Apologies in advance: Peter Stott, Jayashreev Revadekar (JR), Kate Willett (ill) (KW), Tom Peterson (TP), Akiyo Yatagai (AY)

Call in details circulated by email

1. Review of actions from previous meeting

Unresolved on previous call from two calls previous:

ACTION: Richard to contact Ian Jolliffe and find out whether posited Significance submission is still planned to go ahead.

Contacted Ian, but subsequently the editor of Significance contacted REC to discuss the possibility of putting together a collection of articles around the theme of "large data", and wondering whether REC could organise a climate-related contribution to this collection. In statistics at present there are hot topics on massive data volume holdings in different areas of science. Interest in thinking about managing, analyzing, coping with these data sizes. Could it be possible to link into that theme?

JL: We could contribute. It is a problem that we have. Our data holdings are somewhat smaller. Still have problem of how to make data accessible.

REC: IN addition the challenge of analyzing the data appropriately e.g. when creating global data products - computational challenges for huge covariance matrices. The need for automated quality controls is also a "large data set" issue.

NEW ACTION: REC to discuss further with JL, PT and others on potential article for Significance on collection of papers on "large data", as and when plans are firmed up by editor.

ACTION: GS to take forwards BIPM recognition of the initiative.
GS not in attendance and MdP and AM had no further knowledge.

Raised on the last call:

ACTION: TP to provide request for short doc from ISTI detailing what we would like and why in terms of any renegotiation of WMO Res 40.

TP provided request to PT who led production of such a document that was discussed at the CCI management meeting held in Denver.

BT: Can TP update us?

NEW ACTION: TP to send email summary of CCI's position on the WMO Res 40 annexes renegotiation issue.

ACTION: PT to invite Andrea Merlone and Chris Merchant to join as affiliate members.
This was done and they are now added as members.

ACTION: ALL to provide feedback on both the generic poster and powerpoint by next Friday (16th Sept) and then PT will recirculate one last time. Then post to website.
Poster and generic talk now posted to website and generic materials folder added

ACTION: KW / RA to find out what if anything can be done at COP17 to raise

awareness and report back
Neither KW or RA were in attendance to update.
KW: Contacted Peter Stott and waiting for a reply.

2. Regional Report

None forthcoming

3. Updates on overarching initiative related activities

3.1 Meetings held and planned

ISI Dublin meeting - REC

REC gave a presentation on the initiative - geared towards transparency and openness. Quite a lot of interest. REC to provide links to online material in due course.

ACRE meeting - KW / RA

PT characterised as a success

KW: lots and lots of effort towards data collection and rescue. more funding always needed. Work needed to help coordinate all the outputs - ISTI is very useful here.

GCOS Steering Committee - KW

PT again characterised as a success and highlighted some feedback relating to prioritization of data over provenance at this early stage.

KW: Query from some countries on why they should give up their data when they already provide a number of stations to the World Data Centre - a valid point when not all countries (UK!) provide all their data freely.

KW: Multiple calls for databanks and dataportals for various different areas (polar research, terrestrial data etc.). Coordination with other efforts would be good.

KW: GCOS support but anything formal would come through AOPC - report at next meeting.

European Climate Data Management Workshop - KW

Two posters were kindly displayed by Dan Hollis (Met Office Hadley Centre).

WCRP OSC - PT

See <http://surfacetemperatures.blogspot.com/2011/10/wcrp-osc-thoughts.html>

ITS9 -PT/GS

There are a total of 12 climate related talk abstracts plus the ISTI keynote. In total there are six ISTI related talks that will be split across two sessions. Exact mix is still determined although PT has provided a recommendation as to how to split up the climate sessions to the organizers.

Meteomet kick-off meeting - AM

60 attendees from 20 countries. WMO and CIMO also. Good success, TV coverage. ISTI fits into WP4 of the initiative most cleanly. First priority is to collect some new data series - long-lasting series. Will be available for databank. Looking at developing two software packages to analyze the data from a metrologists viewpoint. 3 year project lifetime. Absolutely need help and aim to provide a tool that is useful.

REC: B type uncertainties?

A type uncertainties are statistical uncertainties. B type uncertainty is the instrumental uncertainty. Aim is to develop a calculation model for B Type.

MdP: GUM is the best source : see <http://www.bipm.org/en/publications/guides/gum.html>.

JL: What is the new data you are looking for?

AM: We are looking for series. With metadata. In Italy we have two volunteers getting very very long series. Other will come from other Countries in Europe.

JL: Is very good metadata a requirement?

AM: Yes, questionnaire will be prepared.

PT: Databank has much data and efforts such as HISTALP could be relevant

Other updates - ALL

REC: There is a meeting planned by the Royal Statistical Society next April, maybe with Royal Met. Soc. on climate measurement. Mentioned to KW. 1 day meeting in London, statisticians, meteorologists, metrologists planned. NPL involvement? Mdp: Definitely, possibly.

CM: What is the scope? Oceans and land?

REC: Don't feel prescriptive.

CM and REC to discuss further offline.

3.2 Papers

Paper submission to Environmetrics - REC

Submitted end of Sept. Under review.

BAMS paper status - PT / KW

This has now had the proofs read and will appear with the November edition of the journal. (current issue is September....!)

Paper to metrologia? - GS/Mdp

On hold presently. Will revisit soon.

4. Working Group updates

Annual progress reports were distributed in advance. Call participants will be asked to adopt these as the official versions or to suggest specific modifications after presentations by the chairs.

4.1 Databank - JL

The Databank has now been populated with numerous sources of data at both the daily and monthly timescales during the past year. Daily data are available in Stage 1 and Stage 2 formats for the following.

- Australia
- Brazil
- Channel-islands
- ECA/KNMI
- Ecuador
- ISPD (International Surface Pressure Data)
- IPY

- Swiss
- Sydney
- Tunisia-Morocco
- Japan
- Mexico
- Pitcairnisland
- SACA/KNMI
- Spain
- Uruguay-INIA
- US Forts
- Vietnam

Monthly summary data are available in Stage 1 and Stage 2 formats for the following.

- Antarctica-SCAR-reader
- Antarctica-South Pole
- Arctic
- Australia
- Canada
- Central Asia
- Colonial Era Archive
- East Africa
- GHCN-Monthly version 2
- GHCN-M version 2 source data
- HadCRU version 3
- Histalp

- Japan
- ECA/KNMI
- Russia
- UK Met Office historical
- World Weather Records

Each source dataset is held in its own subdirectory and a README file accompanies the data in the Stage 1 directory, providing basic information on the data provider and the data. An INVENTORY file is included in each source subdirectory within the Stage 2 directory structure. This file provides a list of each station for the particular source and other metadata including station id, name, location, elevation, and first and last year of data.

A limited amount of Stage 3 and Stage 4 data has been added to the Databank on the Daily side - from the Global Historical Climatology Network-Daily dataset - contains more than 25,000 stations with daily maximum and minimum temperature. This is also available for conversion to monthly summaries and future use on the Monthly side of the databank.

Work has now begun on developing methods of data merging and a process for developing and launching the first version of the merged Stage 3 data. The working group has as its goal the launch of version 1 of the global databank in April 2012.

To aid the collection of new source data - a Data Submission Guidance document is now available. This provides instructions for data providers - a description of the process for providing raw native digitized (Stage 1) data to the databank and encourages data providers to follow basic guidelines. A cover letter to accompany the guidance is also available and provides a brief overview of the International Surface Temperature Initiative and the role of the global databank within the wider context of the grand initiative. The cover letter and guidance document are available on the databank website:

<http://www.surface temperatures.org/databank> .

A certificate of appreciation also is under development. This will convey to National Meteorological and Hydrological Services the gratitude of WMO's Commission for Climatology, the Global Climate Observing System, and the World Climate Research Programme for their support of the International Surface Temperature Initiative and broader efforts to meet 21st century needs for climate information.

Objectives for the coming year:

- 1) Continue to add sources of Daily and Monthly timescale data to the Databank. Work with DWG members and others in identifying and collecting readily available sources of digital data.
- 2) Build upon Data Rescue activities and leverage crowd sourcing efforts to begin volunteer digitization of land surface records.
- 3) Complete position paper on version control and provenance available for public comment.
- 4) Develop an approved methodology for merging sources of data to create a monthly Stage 3 data product. Include a hierarchy of source data from which to build the merged dataset.
- 5) Launch version 1 of the Databank for monthly timescale data in April 2012, making all data, processes, and software freely available and accessible.
- 6) Complete and submit journal article describing version 1 of the Databank and its underlying principles.
- 7) Work with the Benchmarking and Assessment working group to expand opportunities for incorporating Stage 4 and 5 data into the databank.

The above and other information is summarized in the Databank Working Group annual status report that Peter circulated on 11/8.

JR noted in apologies that she was working on gaining Indian data for the databank.
BT: This would be a fantastic development.

There is a wiki available for anyone to edit if given a username and password. Jared Rennie is the owner to ask for a username / password.

MdP: Compare to Berkeley effort?

JL: More than 95% of the data is in the current databank, but databank has more. GSOD has some issues in the data which the effort haven't really addressed yet.

PT: Berkeley uses less data than will be available in first databank release.

CM: Is the time window up to present?

JL: Yes, up to present. But also historical. Some data decks are routinely updated.

4.1.1 - provenance and version control - JL

Data provenance methodology was agreed upon and is being used in development of Stage 2 data. It consists of the use of Data Provenance Tracking Flags.

Five (5) Data Provenance Tracking (DPT) flags are assigned to each observation within the Stage 2 data files. These flags provide information on the origins and types of Stage 0 and Stage 1 data. The 5 flags are: (1) Stage 0 Source, (2) Stage 1 Source, (3) Data Type, (4) Mode of Digitization, and (5) Mode of Transmission/Collection. Additional flags can be added as the need arises, and the information contained within each DPT flag can be expanded as necessary to completely define a new type of observation.

This task team recently investigated the potential uses for Unique Identifiers (UIDs), which are now being implemented in the ICOADS marine dataset. While the unique nature of ocean observations creates a need for the use of UIDs the team determined that they would not benefit the land surface databank. Additional information is available at <http://www.surface temperatures.org/databank/provenance-and-version-control-task-team>

Will look to version the entire databank through some x.y.z versioning system. Minor (z) versus moderate (y) versus wholesale (x) changes. Major change may have a paper, moderate a tech note and minor a change in a readme noted. Revision version protocol will be documented in the databank paper.

REC: Useful for users to have guidance online regarding the kinds of updates that would be considered as minor / moderate / major.

JL: Will try to be sure to include that.

4.1.2 - data rescue - PT

Data rescue task team has continued to have calls. Some data has been made available as a result but much work needs to be done. Largely it is a question of resources. It is hoped that we can get resource to support crowdsourcing through the citizen science alliance (see item 5).

Adoption of Databank WG report - sign off by: PT BTMdeP,REC

AMh: IEDRO and ACMAD are working on rescuing the microfiche. AMh is the contact point at ACMAD. Effort made to get NMS's to agree to public.

PT: Strongly welcome this development. This data will then be secure for the future. There was a real risk it would be lost.

4.2 Benchmarking - KW

Kate attended the COST HOME 7th Seminar:

- invited to talk about ISTI and the Benchmarking Programme

- most participants are only interested in regional data and so the global scope was something new but may take some selling
- moving towards a suite of fairly blackbox methods - a number perform well depending on the statistics assessed - good reproduction of regional timeseries but false alarm rate is quite high?
- a small network of benchmarks has been created at the monthly and daily scale and algorithms validated against it - this has shown the enormous value in benchmarking as its the only way to compare these algorithms
- caution over creation of the errors in the benchmarks - users were unhappy with various aspects that were not thought to be very realistic and therefore not very fair
- getting the balance between blind benchmarks and providing a useful research tool is tricky
- we still don't really understand the details of inhomogeneity. Some countries have a good idea of the composite of a large proportions of their inhomogeneities (i.e. 20% shelter changes, 40% station moves etc). That is very useful information as we can be country specific in our addition of errors to the benchmarks in some error worlds. So that is one side of it. The other side of it, are the physical characteristics of those inhomogeneities across the seasonal and even diurnal cycle.
- potential for a further research project on daily benchmarking - ISTI support would be appreciated - this could work alongside the Benchmarking group. The funded project will do more novel research, the benchmarking group will pull the best bits of this together within the benchmarking programme.

Team Creation:

- Robert waiting on Kate to provide model data and station data to play with.

Team Validation:

- plan to use the COST HOME benchmarks to test different validation methods.
- Ian Jolliffe to present a poster at the 5th International Verification Workshop (<http://www.cawcr.gov.au/events/verif2011/>), Melbourne, Australia, December 2011

Not much progress over the last few months. Time to pick up the pace.

Need to work with Databank team on storage for the 8-10 analog worlds of station data as these will be huge. Can the gosc portal hold this?

Progress Report written, accepted by the Benchmarking group.

AM: Propose that meteomet are working on software tool, could they be added to WG?

ACTION: KW to discuss with AM membership from meteomet on the benchmarking WG and report back.

Adoption of Benchmarking and assessment WG report: Sign off by: PT BT, MdeP, AM, REC

5. Funding and opportunities discussion

PT was unsuccessful in getting forward a large NSF funding bid.

JL and PT have led a bid for internal funding support for crowdsourcing land data digitisation.

AM: Two potential opportunities.

1. Research mobility grants within meteomet initiative. Could help in some cross-linking activity. For people from Eastern European countries. Currently no candidates under WP4.

2. In 2013 the EMRP will have a second call in on environment. Could add proposals of topic to be submitted. Discussion on proposal should start in 2012, in order a common (ISTI-Metrology community) proposal can be submitted at stage 1 of the call (expected feb 2013). Then leads to potential funded JRP.

6. Meteomet collaboration agreement - AM / PT

Meteomet have requested a formal memorandum of understanding. This will allow support from NMIs in Europe and formal meteomet engagement. PT will put together a draft in early December once IPCC commitments have reduced and circulate the relevant materials for comment prior to submission. There

may also be some funding available although a European NMI lead would be required to access these funds.

ACTION: PT and AM to iterate and circulate a memorandum of understanding document between the initiative and Meteomet for approval by steering committee by 12/15

7. Research Network on Surface Temperatures update - CM / PT

There has been an initial call of this effort's steering committee which PT sits on and is led by CM. Presently the focus is on an initial meeting to be held in 25 to 27 June 2012 in Edinburgh. The effort is looking at all aspects of surface temperatures not just the in-situ instrumental land component so is much broader than our effort. Its remit is to bring researchers together to address issues - there is no hard funding to undertake new research efforts per se although a minor bit of postdoc support for a postdoc under CM does exist.

Visiting scientist facility exists. Must be cross-domain e.g. LSAT and Satellite surface temperatures.

8. AOB

Propose next call on Jan 11th at 8am EST. Note that we need to produce a progress report by the end of that month.