

Hello and a Happy New Year to all Databank Subgroup members!

With the start of the new year I thought it would be a good time to update everyone on how things have progressed since our first conference call on October 29. Progress continues to be made, and in some areas faster than others. Below I have included the action items from the conference call (in italics) with the status of each item in bold font. I am also attaching a paper about this grand challenge that has been submitted to BAMS. Peter has given the ok for me to send this, but Please DO NOT cite or distribute until published. Thanks to everyone for your continued interest and participation. Please let me know if any questions.

Jay

Membership

Additional members need to be added in key functional areas and regions of the world.

To support data rescue activities Rob Allan (ACRE) will receive an invitation to join.

Rob Allan has accepted an invitation from Peter Thorne to provide expertise on data rescue.

Need more representation from S. America to address data provision outside Argentina. Steve Worley to provide potential member from Brazil by 10 November. Barbara or Matilde can provide potential member from Uruguay by 10 November.

Waldenio Almeida (CPTEC/Brazil) has volunteered for the group and awaiting management approval.

Need representation from Africa. Several names put forward. Steve Worley and Jay L. will continue to pursue and resolve by 15 November. John Christy also offered to serve as representative given his experience there.

Still pending resolution.

Databank Development

Initial offer to use GHCN-D and GHCN-M datasets as foundation of databank was put on hold in favor of small test datasets that can be established from Stage 0 through Stage 3. NCDC to pursue and establish a process for creating these test data sets before end of November. Goal to have test datasets available by end of December.

NCDC has a small set of Stage 0 through Stage 3 data for stations from the U.S. which will provide an example of how the full databank will be structured. These examples will be made available from the World Data Center A website. The site is currently being constructed.

Efforts to build on current global data holdings continue as described below.

JMA, via Koji Ishihara, is in the process of providing Japan's historical record of monthly and daily max/min/mean temperature data following a formal request from NCDC to the Director General of JMA's Observation Department.

Manola Brunet, Director of the Centre for Climate Change at Univ Rovira i Virgili has offered to provide max/min temperature data for 22 long-term daily stations developed some time ago as part of the EMULATE project. These data are fully traceable and will include imaged, raw, and adjusted data.

Waldenio Almeida indicates that CPTEC/INPE has recovered and reprocessed their 15-year GTS database. CPTEC also has data from several AWS Brazilian regional networks that are not in GTS. At the Brazilian Met. Service, work is being done to digitize data from observation books. These data from the thousands of volumes has the potential to greatly increase spatial and temporal coverage in Brazil.

Mirroring is not an immediate priority but one site has been offered by Steve Worley at NCAR.

From Waldenio Almeida - The Databank project is in line with INPE's initiatives and way. INPE (and maybe also the Brazilian Met Service) likely would be interested in being a database mirror site.

Data will reside in ASCII format with NetCDF a high priority, especially as data of higher temporal resolution included. Relational Databases appropriate for metadata and to improve data access and visualization (to be addressed by Data Access and Visualization Team).

Still pending establishment of a Data Access and Visualization Team.

Data Provenance and Version Control

This is of highest priority. Processes and guidance need to be established early on.

This will be initiated by a Working Team: Initial contributions by Steve Worley, John Christy, and Jeremy Tandy. Jay Lawrimore to coordinate activity and address other resources at NCDC. To establish conference call for latter half of November to discuss direction of this working team.

Steve Worley to provide starting point by pulling from ICOADS data provenance and version control library of documents. By 10 November.

Steve Worley has provided a starting document for this activity. NCDC's Debra Braun has been brought onto this working team. She is a senior analyst and in the past was involved with COADS, so she has experience with global surface obs collection and management activities including digitization oversight. Kick-off activities for this working group will take place in January.

Data Rescue and Crowdsourcing

Data rescue and crowdsourcing will be an essential aspect of this effort and needs to receive considerable attention. Millions of images currently exist – major efforts over the past decade including NCDC's Climate Database Modernization Program have provided rich sources.

Peter Thorne along with Rod Hutchinson has formed a Data Rescue Task Team comprised of several international experts. Also, NCDC had further discussions with Chris Lintott from Galaxy Zoo and set up a visit to NCDC for Chris from 15-16 December. This visit,

which will provide the necessary background for Chris to establish a crowdsourcing plan for the millions of already imaged forms for land surface data, had to be postponed at the last minute because of a personal injury. Will be rescheduled for first quarter of 2011.

Data Inventory

As a starting point, this Subgroup needs to develop an inventory of the data that currently exists.

This will begin with an inventory of GHCN-Monthly and GHCN-Daily data. Jay and Matt Menne will coordinate at NCDC, and Rod Hutchinson also has an inventory of data from Pacific that can be provided. Inventory to be provided by 3 December.

Inventory of data has been developed and will be placed on the World Data Center-A website when it is up and running.

Providing an inventory of data in imaged format much more difficult and inventory currently limited. To the extent that an inventory of imaged data exists, it will be provided. Peter Thorne to coordinate. Timeline TBD.

Still in progress. Looking for additional resources to support this task.

Near real-time and historical data via WMO processes

Subgroup recognizes the need to address issues requiring WMO action (annual World Weather Record transition from decadal to annual, daily climate summaries via daily climate bulletins instead of synoptic, and inclusion of metadata in CLIMAT bulletins). Jeremy Tandy to be involved via Expert Team on Metadata and Data Interoperability that he chairs. Jay L. also to coordinate with Stephan Bojinski at WMO regarding Expert Team on Data Representation and Codes and potential coordination already taken place. By 5 November.

Jay has held discussions with the lead for the World Weather Records 2001-2010 initiative in an effort to send inquiries to countries regarding their opinion on a transition from decadal to annual updates. Also discussed with Stephan Bojinski regarding the expert

team's support for this effort. Identified need to develop requirements associated with need for daily climate bulletins.

Building Momentum

Following progress by this Subgroup's Working Teams (Data Provenance and Version Control; Data Rescue and Crowd Sourcing) this subgroup will address and plan a workshop to include this subgroups members and others who are key to this effort.

Under ongoing discussions.

These meeting notes to be placed online on surface temperature .org site by Peter Thorne by 3 November.

Meeting notes are available online at <http://www.surface temperatures.org/databank> .

Jay Lawrimore to begin developing Databank Scoping Document with executive summary for management. Date TBD.

An overview of the project for general audiences has been submitted to BAMS for publication. Once published I think everyone will find it helpful in providing a good overview of the project as a whole. I am providing a draft copy with this e-mail for everyone's information. Please DO NOT CIRCULATE OR CITE.