

## **Databank Working Group**

2013 Progress Report

January 2014

### **Current Members:**

John Christy	- University of Alabama, Huntsville, USA
Meaghan Flannery	- Australia Bureau of Meteorology
Waldenio Gambi de Almeida	- CPTEC/INPE, Brazil
Kenji Kamiguchi	- Japan Meteorological Agency
Albert Klein-Tank	- KNMI, Netherlands
Jay Lawrimore (Chair)	- NOAA NCDC, USA
David Lister	- Climatic Research Unit, East Anglia, UK
Matthew Menne	- NOAA NCDC, USA
Albert Mhanda	- ACMAD, Niger
Colin Morice	- UK Met Office, Exeter, UK
Vyacheslav Razuvaev	- Roshydromet, Russia
Jared Rennie	- CICS-NC/NOAA NCDC, USA
Madeleine Renom	- IFFC, Univ of the Republic, Uruguay
Matilde Rusticucci	- Univ of Buenos Aires, Argentina
Jeremy Tandy	- UK Met Office, Exeter, UK
Peter Thorne (ex-officio)	- NERSC, Bergen, Norway
Steve Worley	- National Center for Atmospheric Research, USA

### **New Members:**

Byron Gleason	- NOAA NCDC, USA
---------------	------------------

### **October 2012 to October 2013 Objectives:**

- 1) Incorporate user feedback, making further improvements to the monthly Stage-3 dataset.
- 2) Complete a full version 1 release in January 2013.
- 3) Continue to collect and add new sources of monthly, daily, and sub-daily data to the databank.
- 4) Complete incremental updates to the monthly Stage-3 dataset.
- 5) Begin development of the daily Stage-3 dataset.
- 6) Work with the Benchmarking and Assessment working group to expand opportunities for incorporating Stage 4 and 5 data into the databank.
- 7) Complete publication of articles describing the Databank and its goals as part of the International Surface Temperature Initiative.

Objectives added for 2013 reporting period in the Implementation Plan refresh:

- 1) Creation of netcdf files for the merged holdings in July 2013
- 2) Definition of a format for the parallel measurements database, October 2013

## Objectives Met:

- 1) Incorporate user feedback, making further improvements to the monthly Stage-3 dataset

The Stage 3 databank was initially released in beta form in October 2012. A recommended merge along with seven alternative merged datasets (variants) were provided on the ISTI website. The merging methodology is described in a document made available in early October 2012 at [ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/merging\\_methodology.pdf](ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/merging_methodology.pdf). The availability of the stage 3 data to the user community resulted in comments and recommendations for improvements which resulted in a second beta release in January 2013. Use of the databank continued with additional improvements subsequently leading to a third and a fourth beta release in May 2013 and October 2013, respectively. A description of the changes incorporated into each beta release is available at [ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/updates\\_to\\_beta\[2-4\].pdf](ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/updates_to_beta[2-4].pdf).

- 2) Continue to collect and add new sources of monthly, daily, and sub-daily data to the databank.

Two new sources of monthly data (mcdw-unpublished, preliminary CLIMAT) and three new sources of daily data (HadISD, ECAD, and SACAD non-blended) were added to Stage 1 of the databank during the past year.

- 3) Complete incremental updates to the databank.

Updates to the databank continued through the past year with four beta releases as described in section 1 above.

- 4) Begin development of the daily Stage-3 dataset.

Collection of daily data continues and development of the Stage-3 dataset is being accomplished through a parallel effort involving NCDC's Global Historical Climatology Network-Daily dataset. This dataset contains more than 25,000 stations with daily maximum and minimum temperature, snow observations, and more than 80,000 stations with daily total precipitation. The earliest observations are from the 1800s with the majority of stations having data from the middle of the 20<sup>th</sup> century to present. The dataset is updated each day with thousands of observations made the previous day. It is available at <http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/>.

- 5) Work with the Benchmarking and Assessment working group to expand opportunities for incorporating Stage 4 and 5 data into the databank.

This has been partially met. Two members of the Databank working group remained closely involved with the Benchmarking and Assessment working group (BAWG) during the past year. Progress was made and several objectives were at least

partially met by the BAWG including drafting a concepts paper and developing software to produce analog clean and error worlds. Progress also was made on designing methods for assessing the results of tests. No Stage 4 or 5 data are yet available for addition to the Databank. The Databank WG will remain involved with the BAWG and will be prepared to add Stage 4 and 5 data as further progress is made.

- 6) Complete publication of articles describing the Databank and its goals as part of the International Surface Temperature Initiative.

This has been partially met. A short article outlining the databank was published in EOS in early 2013. Presentations also were made at the Annual Meeting of the American Meteorological Society in January 2013 and at the Fall meeting of the American Geophysical Union in December 2013. Publication of the full technical details associated with data collection and the merge process associated with creation of the Stage-3 databank was delayed. Although the article was completed and accepted by the Geosciences Data Journal, publication was postponed so that it will coincide with the version 1 release of the databank.

- 7) creation of netcdf format holdings

CF compliant netcdf files were created in the summer of 2013 and included in the beta 4 release.

#### **Objectives Not Met:**

- 1) Complete a full version 1 release in January 2013.

Release of version 1 of the databank (Stage 3 merged) was delayed due to new 11<sup>th</sup>-hour requirements imposed by NCDC management. These new requirements include a code review of all software responsible for source reformatting, calculation of monthly means from daily observations, and algorithms that conduct the source merge. The code review identified several enhancements that can be made to improve the long-term maintenance of the software. These enhancements will be made in the coming weeks.

- 2) Definition of a format for the parallel measurements database

It was deemed that taking necessary steps to formally release version 1 of the merged holdings took precedence. It is planned that this will be addressed in the next reporting period.

#### **2013 Annual Overview:**

Good progress was made in 2013 toward release of version 1 of the Stage-3 databank. Feedback was received from the user community and additional review and analysis led to improvements in the merged dataset. A total of four beta

releases were made available with each successive release improving on the previous. The addition of five new sources of monthly and daily data further improved the databank, bringing the total number of sources to 58. Although the goal of releasing version 1 was not achieved as expected, the additional steps which will be taken before release will be beneficial to the long-term success of the effort. For example the code review of the databank processing software led to recommendations for changes that will make the software suite easier to maintain in the coming years. Once all release requirements are met, version 1 will be made available and accompanied by web and media announcements.

**Objectives for October 2013 to October 2014:**

- 1) Complete software enhancements associated with code review
- 2) Complete operational readiness requirements for version 1 release
- 3) Version 1 release of Stage-3 data
- 4) Begin collection of parallel measurements
- 5) Enhance metadata collections
- 6) Add to collections in data sparse areas

**Suggested timeline and plan for achieving objectives:**

<b>Objective</b>	<b>Description</b>	<b>Responsible Members</b>	<b>Deadline</b>
Advancing exchange of daily climate summaries on a routine basis	Completion of development and testing of new CLIMAT template containing daily observations	NCDC	December 2013
Completion of software enhancements associated with code review	Improve long-term maintainability of stage 1 through 3 software	Rennie, Gleason	January 2014
Completion of operational readiness requirements for version 1 release	Meeting documentation and review requirements internal to NCDC	Lawrimore	February 2014
Version 1 release of Stage-3 data	Release version 1 of the Stage-3 dataset with variants. Web and media announcements.	All, lead by Jay Lawrimore	March 2014
Metadata collection strategy	Documentation of Working Group's strategy to pursue metadata holdings for existing data holdings	Databank Working Group	July 2014
Work with ACRE / ICOADS etc. to investigate potential extension of IMMA format to land meteorological data	Improve long-term archive of land surface data	Lawrimore, Woodruff (Guest expert)	August 2014
Implement NRT and period	Provide continuing	Rennie	September

of record updates to the databank and document how these are done	updates to Databank		2014
Begin collection of parallel measurements	Develop formatting structure for storing parallel measurements and begin collection (At least 3 sets of parallel measurements collected.)	All, lead by Jared Rennie and Victor Venema	September 2014
Add to collections in data sparse areas	Enhance data collections in data sparse areas, principally Africa and S. America. Work with ACRE and EarthTemp Network.	All	September 2014
Advancing daily databank activities	Completion and publication on website of plan for developing and operating daily databank	Databank working group	October 2014
Begin extension of databank to multi-elemental holdings by creating infrastructure for precipitation and/or snow depth	Databank expansion	All, lead by Rennie	March 2015
<b>Ongoing activities</b>			
Advocacy of the databank, efforts to augment holdings	Every effort should be made to engender data submissions	Steering committee, Databank working group	Ongoing
Data rescue	Continued pursuit of funding proposal for support of crowdsourcing of already imaged forms (such as NOAA foreign data library)	Data rescue task team / Databank Working Group	Ongoing until successful
Parallel measurements database data collection	Pursuit of parallel measurements data holdings	Databank Working Group / Victor Venema	Continuous

Note that a number of these suggested activities are delayed relative to the Implementation Plan. This reflects the need to prioritize in the short-term taking necessary steps to enable release of a first version of the databank holdings.