# Responding to the Need for Better Global Temperature Data

PAGES 61–62

Since the 1990s a number of national institutions have developed and maintained global data sets of land surface air temperature [*Peterson and Vose*, 1997; *Hansen et al.*, 2010; *Jones et al.*, 2012; *Menne et al.*, 2012]. These efforts have led to great advances in understanding how Earth's temperatures have varied and changed. They also serve as essential sources of a fundamental climate variable, which is crucial for interpreting climate evolution in response to the interplay of radiative forcing, climate feedbacks, and ocean heating [e.g., *Hansen et al.*, 2011]. However, more can be done to improve global surface temperature collections while enhancing data management, access, and public transparency with which data are collected, processed, and converted into climate information. To address these needs, the International Surface Temperature Initiative (ISTI), which began through a partnership of scientists from around the world [*Thorne et al.*, 2011], released its first beta version of a global land surface databank in October 2012.

One of the key improvements over previous data holdings is greater spatial coverage of temperature observations from the 1800s to the present. Together with other initiatives, the databank has benefited from a renewed focus on rescuing, collecting, and integrating data from original forms to create the most complete data collection possible. This is only a start, as it is likely that as much data exist as images or hard copy records created prior to 1950 as are presently available in digital records.

Data provenance also has been improved, which is necessary for more fully characterizing uncertainties inherent in the climate record. By collecting data in the most original form possible and by documenting and making available information on each observation, all of the factors that influence the temperature record (e.g., method of measurement and computation, observing instrumentation and practices, and quality control) can be better accounted for, and uncertainties in the temperature record can be more fully quantified. In addition, more attention to version control and open access to all software used throughout the databank development process are being provided.

The databank is being constructed and made available in stages (Figure 1). Stage 0 consists of observations in their original form, whether recorded on paper and housed in various archives or converted to photographic or scanned images. Stage 1 contains digital data in their native format such as ASCII text, spreadsheet, or other electronic document. This reduces the possibility that errors could occur during translation while benefiting the provider by not requiring extra effort on his or her part. Ideally, no changes to the original observations should be applied prior to submission to the databank so that the provenance can be better ensured leading up to and through the point where efforts at quality control and homogenization of the databank are to be accomplished. Homogenization is essential to climate assessment because it removes artifacts in the temperature record associated with nonclimatic influences such as changes that frequently occur in observer practices, instrumentation, and station environment.

Once collected, all data are converted to a common format in stage 2, and an inventory is produced containing available metadata. At a minimum, this typically consists of a station identifier, name, latitude, longitude, elevation, and beginning and ending year of data. This step also includes the addition of data provenance tracking flags to help users understand the history of each observation, with information such as the data source, location of original data archive, method and source of digitization, and mode of transmission.

The many stage 2 sources are merged into a single stage 3 data set. This is fraught with complexities associated with the nature of weather and climate data, which are collected by hundreds of thousands of observers in hundreds of countries often using differing languages, observing methods, averaging, and documenting and archive procedures. Because different sources may contain records for the same station (potentially containing differing values), it is necessary to create a process for identifying and removing duplicate stations, merging some sources to produce a more complete station record, and, in some cases, determining when a new station should be added.

From the merged stage 3 data, for which there are presently more than 30,000 unique stations, a suite of quality-controlled (stage 4)

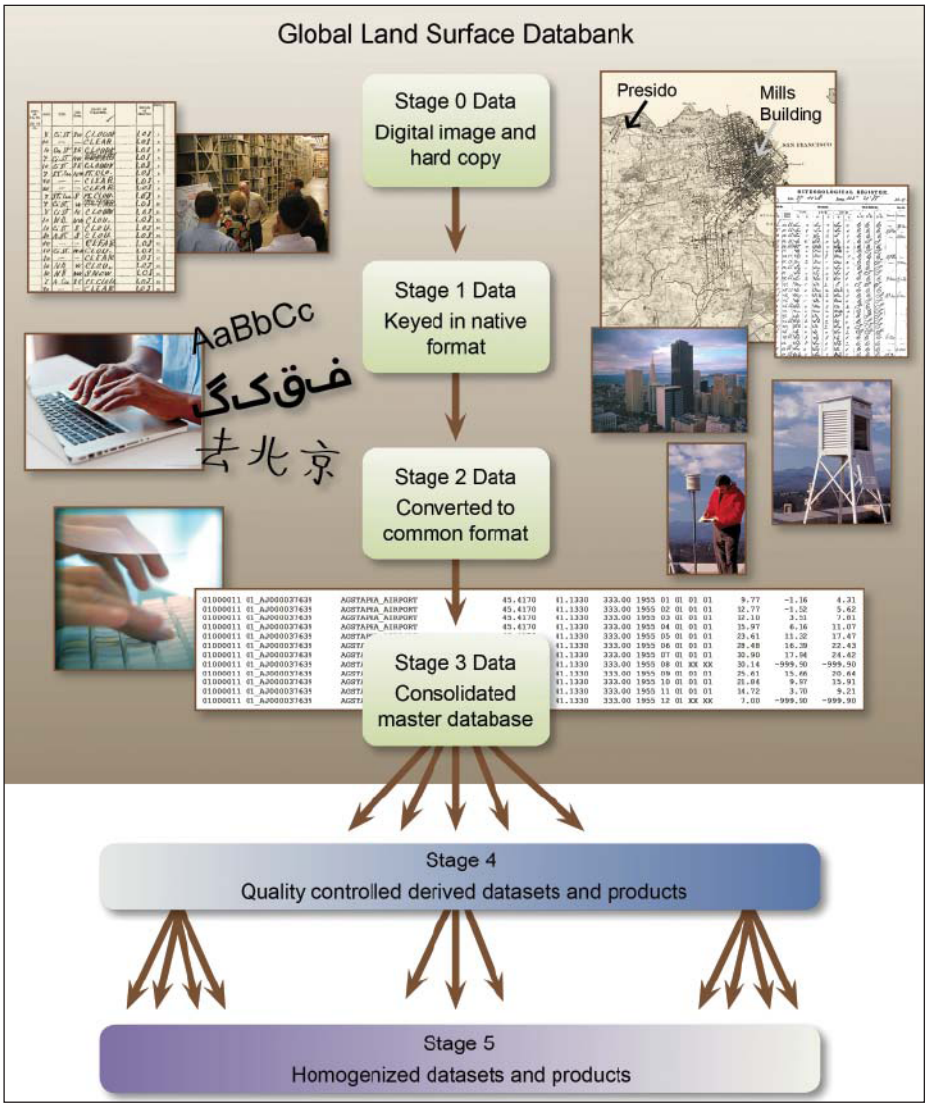


*Fig. 1. The stages of the global land surface databank. Multiple stage 4 and 5 data products can be developed by independent groups.*

and homogenous (stage 5) data sets can be developed. Multiple independent analyses of these data by outside groups are strongly encouraged. These efforts will be evaluated using processes established by the ISTI benchmarking working group [*Thorne et al*., 2011].

Databank submission procedures were designed to make the process of providing data easy while ensuring that submitted data are traceable and of the highest quality possible. All data are available from ftp sites hosted by the Global Observing Systems Information Center (GOSIC; http://gosic.org) and World Data Center A at the National Oceanic and Atmospheric Administration's National Climatic Data Center and at World Data Center B at Obninsk, Russia.

Construction of a land surface databank is a major undertaking requiring time and international coordination, reflected by broad membership in the databank working group. It provides the foundation from which new methods of analysis, assessment of uncertainties, and service to end users can be established and comes at a time when the need for high-quality, traceable, and complete data is greater than ever. While the initial focus is on temperature data on the daily and monthly time scales, other elements and time scales will be added in the future (e.g., precipitation from monthly to hourly scales). Further information is available at http://www.surfacetemperatures.org/databank.

## *References*

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.

Hansen, J., M. Sato, P. Kharecha, and K. von Schuckmann (2011), Earth's energy imbalance and implications, *Atmos. Chem. Phys.*, *11*, 13,421–13,449, doi:10.5194/acp-11-13421-2011.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012), Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010, *J. Geophys. Res.*, *117*, D05127, doi:10.1029/2011JD017139.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston (2012), An overview of the Global Historical Climatology Network-Daily database, *J. Atmos. Oceanic Technol.*, *29*, 897–910, doi:10.1175/JTECH-D-11-00103.1.

Peterson, T. C., and R. S. Vose (1997), An overview of the Global Historical Climatology Network temperature database, *Bull. Am. Meteorol. Soc.*, *78*, 2837–2849.

Thorne, P. W., et al. (2011), Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science, *Bull. Am. Meteorol. Soc.*, *92*, ES40–ES47, doi:10.1175/2011BAMS3124.1.

—Jay Lawrimore, National Climatic Data Center, National Oceanic and Atmospheric Administration (NOAA), Asheville, N. C.; E-mail: Jay.Lawrimore@noaa.gov; and Jared Rennie and Peter Thorne, Cooperative Institute for Climate and Satellites, National Climatic Data Center, NOAA, Asheville, N. C.