Overview
Observational data form the foundation for understanding weather and climate, its variation and change, and its influence on societies, economic development, and natural resources.

As part of the International Surface Temperature Initiative, a renewed effort is underway to create a comprehensive international data holding — a Global Land Surface Databank — of all land meteorological data beginning with temperature.

The purpose is to combine existing holdings with new sources of data and begin the digitization of data not currently digitized to address gaps including inadequate temporal and spatial coverage.

This effort provides information that ensures the integrity of observations from the point of measurement through processing, archiving, and distribution.

Databank Structure
The Databank consists of 4 fundamental stages of data:

Stage 0: Original paper record or digital image of paper/microfiche/microfilm record
Stage 1: Originally keyed data in the native format of the data provider
Stage 2: All data converted to a common format; Data Provenance flags added
Stage 3: Data combined into a single integrated dataset with duplicate data source reconciled
Stage 4: Quality-controlled data
Stage 5: Bias-corrected data

These activities are established as part of the Benchmarking and Assessments effort. It is expected that there will be many different methodologies used in developing unique instances of Stage 4 and Stage 5 data.

Got a Lead to a New Data Source?
Please email data.submission@surfacetemperatures.org and we will help to coordinate with others the most effective way to rescue the data and ensure its addition to the databank.

www.surfacetemperatures.org
data.submission@surfacetemperatures.org

Building the Databank
The Databank initially consists of all available sources of land surface temperature observations on Monthly and Daily timescales (e.g., Monthly and Daily Mean Temperature)

The foundation of the Databank consists of large datasets previously compiled by government organizations:

• NOAA’s Global Historical Climatology Network, Monthly and Daily datasets
• CRUTEM4
• European Climate Assessment (ECA) dataset

Many contributions from Universities and National Meteorological and Hydrological Services are being made, e.g:

• Japan Meteorological Office
• Rovira I Virgili University, Tarragona, Spain
• Royal Netherlands Meteorological Institute

The data will consist of sources that the provider agrees can be released for global distribution free of charge and without restriction.

Stage 3 Beta Release
A beta release of the Stage 3 merged dataset was made in early October 2012. It contains more than 39,000 unique station records which greatly increases the number of stations when compared to other well-known global datasets such as Global Historical Climatology Network-Monthly version 3. The greater spatial density in the Stage 3 dataset results in differences in annual global average temperatures based on raw unadjusted data.

Quality controlled (Stage 4) and homogeneity corrected data holdings (Stage 5) will be developed in the coming year in cooperation with the Benchmarking working group.

Data Provenance
To provide traceability from the Stage 3 merged dataset back to the original Stage 0 data or the earliest stage that exists pro data provenance flags are added at Stage 2.

Provides information on the history of each observation and helps guarantee data authenticity, integrity, and quality — an essential aspect of ensuring confidence in the observational record

Data Provenance Tracking Flags are assigned to every observation. These flags provide information regarding:

• Original data provider for the Stage 0 and Stage 1 data
• Type of data provided (raw, quality-controlled, bias-corrected)
• Mode of calculation (e.g., original observation, from sub-daily data)
• Mode of digitization (e.g., keyed by a named project, auto collected)
• Mode of transmission (e.g., FTP, NMHS Web Service, GTS, Mail, E-mail)
• New data provenance tracking flags can be added as needed

Data Rescue
Data rescue is a vital part of building a complete Databank. There exist very many land meteorological records that are available only in hard copy or image form. These records cannot be readily used in a myriad of scientific contexts without first converting into a digital format and making them easily available through one or more recognized repositories.

Rescue sources:
• Prior to the 1960s many more non-digitized records than digitized
• Even after the 1960s not all data are digitized and readily available
• Data for many regions that are currently poorly sampled in global data records
• Data in many cases are rotting away and seen by the owners as a nuisance. If we do not rescue the data now they will be gone forever.

Data Accessibility
A major goal of the Databank is to provide maximum scientific, economic, and societal benefit by making as much data as possible easily and readily available

Source data are already available on two FTP sites:
or: http://ftp.meteo.ru/pub/data/global_databank/

These sites are accessible from the Global Observing Systems Information Center (GOSIC):
http://gosic.org/GLOBAL_SURFACE_DATABANK/GBO.html
and the International Surface Temperature Initiative:
http://www.surfacetemperatures.org/databank

The International Surface Temperature Initiative’s Global Land Surface Databank and the Importance of Data Rescue