Sophisticated Metadata in Database System for Climatology, Meteorology and Environmental Sciences

Juraj Bartok, Ivana Bartoková, Monika Kľocová, Dušan Mazúrek

MicroStep-MIS, Čavojského 1, Bratislava, Slovakia

monikak@microstep-mis.com / Phone: +421 2 602 00 170

Abstract

As stated by the WMO, the details and history of local conditions, instruments, operating procedures, data processing algorithms and other factors pertinent to interpreting data (i.e. metadata) should be documented and treated with the same care as the data themselves. Complex data management solution developed mainly for climatology but used also in operative meteorology/hydrology, aeronautical climatology, radioactivity monitoring and crisis management is presented. System based on Oracle database works on multiple platforms and offers several cold or hot backup possibilities. Possible data formats this system can decode and ingest into the database are discussed (data from: automatic weather stations, airport weather systems, alphanumerical WMO codes or BUFR, manual key entry, old database systems e.g. CLICOM, spreadsheets; special data as radar, satellite, GRIB). Data processing includes quality control with range, consistency, temporal, spatial checks completed with more complex checks in PL/SQL language and station communication/performance statistics. Accent is given on metadata structures containing descriptive textual or graphical information. Sensor Warehouse contains searchable database entries for sensor parameters traceable to individual sensor piece with assigned serial number. Textual and numerical information is accompanied by copies of datasheets, photos of sensor and environment and calibration certificates with possible e-mail alerts announcing recalibration need. User can add his own sensor or environment properties into the database. Station Observation describes observational programme, siting and environment completed by panorama photos, maps. Full history of instrumentation used for individual weather parameters is tracked. Maintenance Module keeps records of regular and special maintenance. Interfaces are fully web based which creates possibility to access the system in the field via mobile networks. System can ingest all kinds of metadata: WMO required and recommended, nationally developed, special metadata of individual scientific projects and intercomparisons, Leroy siting classification. Several examples of existing system applications are presented.

IMS CLDB database system overview

The main purpose of the IMS CLDB system is to provide:

- data storage management for different environmental data
- storage for meta-data
- interfaces to receive data from different communication systems
- pre-process the data, perform quality control and unify the structure
- user services for data provision, distribution and publication
- user interface for the system is web based application

- system implements common procedures for the data management
- handling of related environmental data and products in order to facilitate exchange and meet individual requirements efficiently

Schematic diagram explains the data flow in the system together with main software modules and its functionality:
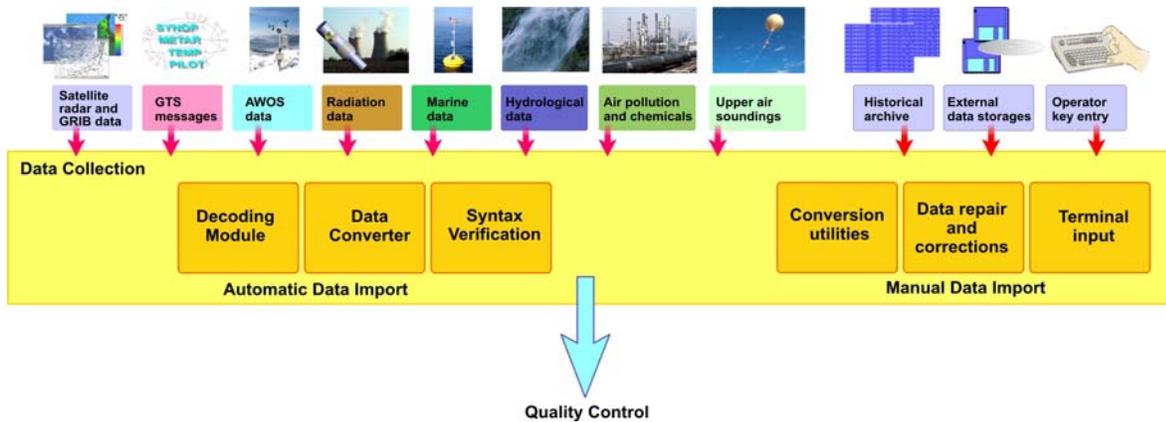
Quality control module checks the data validity, wide variety of checks have been developed:

- elimination of duplicates
- internal consistency checks
- comparison with a global ranges
- comparisons with other observations
- manual monitoring and corrections

Database provides storage of the observed data together with its metadata structures:
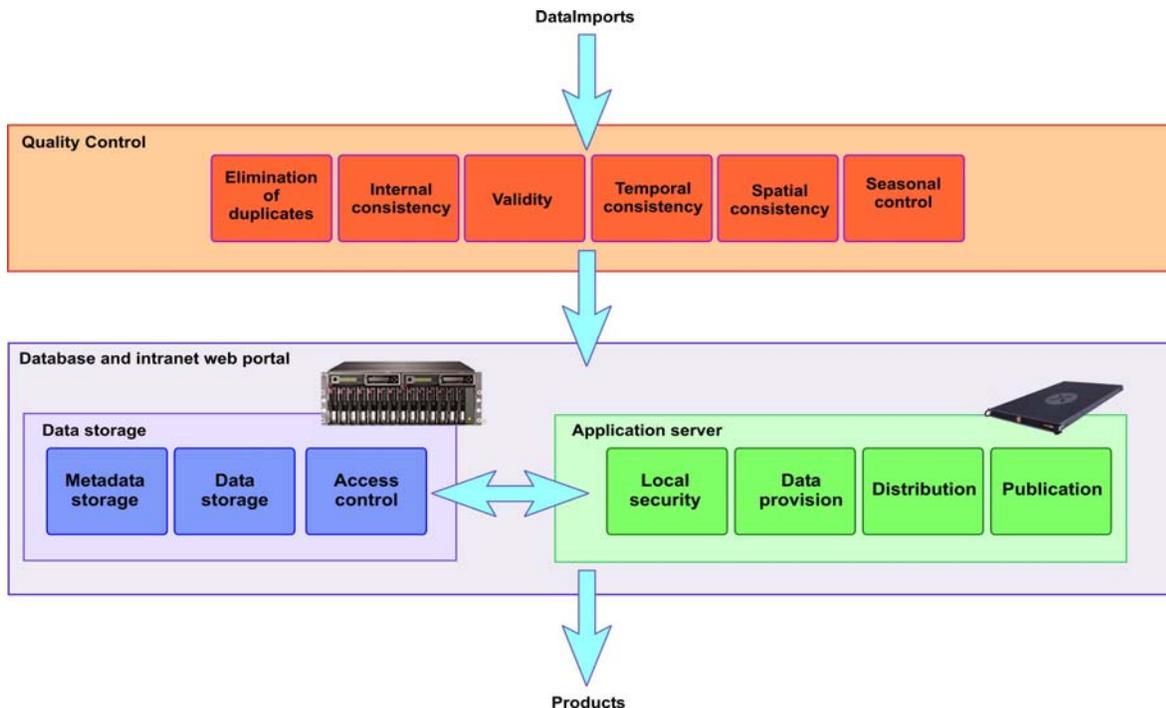
User products are the last step in the data flow:

- The number of types of products varies on the data inputs
- Tabular reports and charts of various environmental elements, analysis of measured environmental data, upper air tables
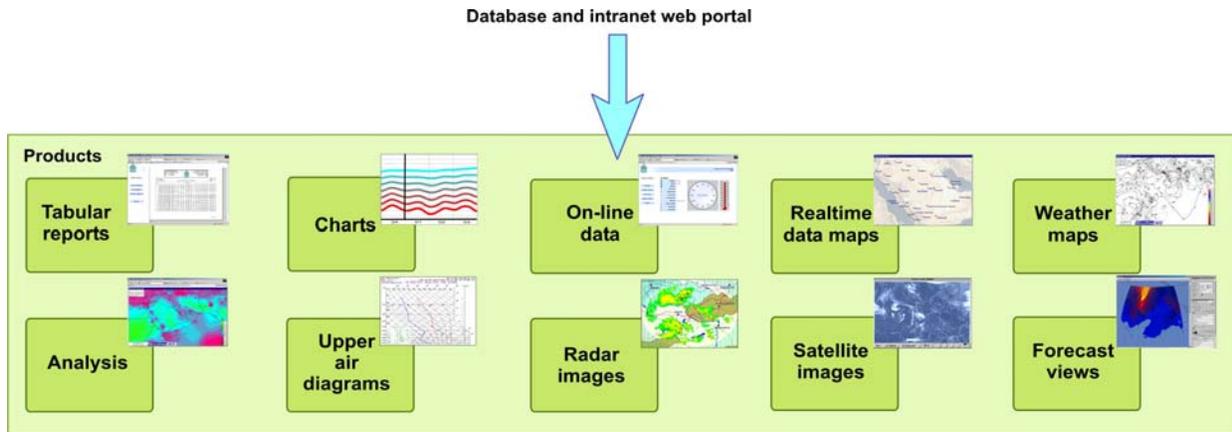

Metadata

The metadata are inevitable part of climate data itself. The CLDB allows storing various meta-data inside the database. Basic Metadata are organized in three groups - station geography, observation description and element description.

Station properties are not only a fixed list. The network manager can add descriptive items via **Property Editor**. For example when a country introduces Leroy classification, its parameters are added into the system by user. Some examples of station parameters are listed below:

- Regional identification: Region ID, Region Name
- Country identification: Country Area, Country Code
- WMO (WMO Index) and ICAO (CCCC) identifiers
- ID number of the station AWS: AWSID
- Station Name
- Time Zone of the station. Can be UTC, UTC+1, UTC+2...
- Station geographic location (Latitude, Longitude, Elevation)
- Quality checking: QC Enabled
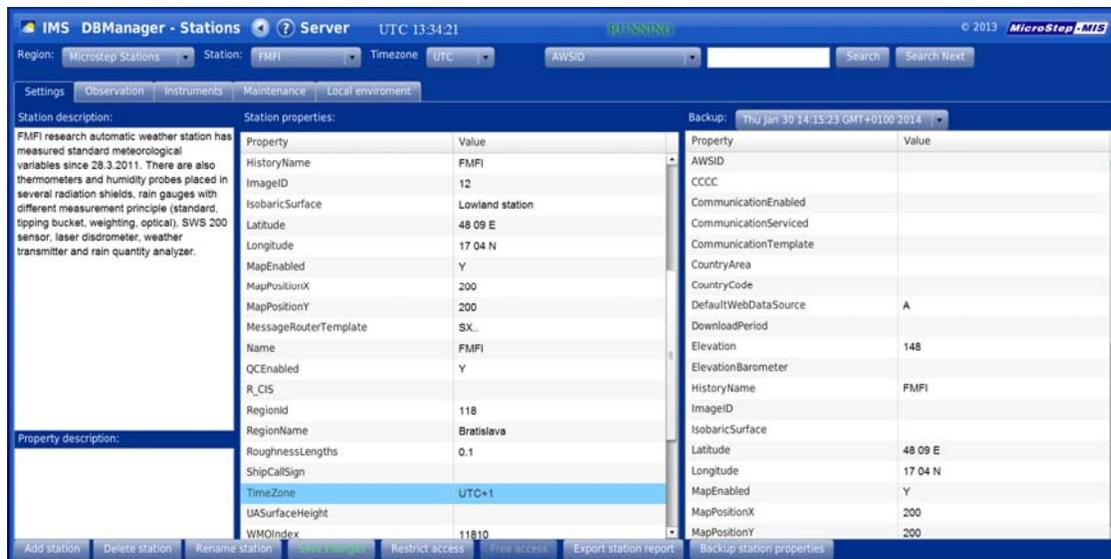
Each variable observed at the station has the following attributes:

- **Start** – Beginning date of observation of particular meteorological element
- **End** – Ending date of observation
- **Instrument** – measuring instrument (sensor) type. These instruments can be described in **Instruments configuration** table.
- **Instrument Serial** – serial number of the measuring instrument
- **Height** – Height of instrument relative to the station ground (e.g. 2m for standard temperature measurements; 10m for wind measurements, ...)
- **Shelter** – type of the device shelter
- **Schema** – observational schema: *Regular* or *Irregular*
- **Interval** – interval of observations in regular schema: 15min, 1hour, 3hours, etc…
- **Sampling** – sampling span
- **Validity** – validity period
- **Description** – short description of the observed element, free text field

Maintenance tab helps to keep the station maintenance records in order.

Local Environment is used to note the metadata of the local station environment which are useful in long-term data processing and its analyzing. Textual descriptions of exposure, soil type and other site conditions are accompanied by photographs, maps, scans of site plans, skyline diagrams:
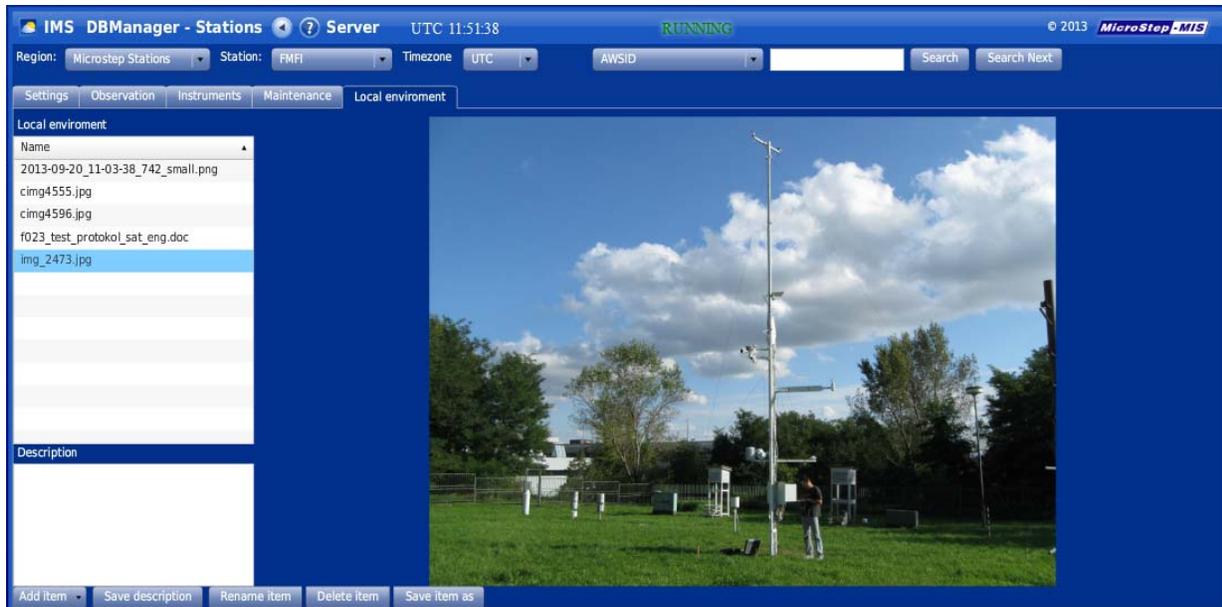
The **Instruments Warehouse** tool deals with measuring instruments. Firstly, sensors are catalogued according to type. To each type, searchable parameters are stored as database entries. Detailed documentation in standard formats such as .pdf or .doc is also stored directly in the database. Specific documentation to each sensor piece with serial number follows, most importantly calibration certificates, notes from in-the-field checks, photos of installation details:


Based on metadata information and responsible persons module in database, e-mails and SMS are sent. For example, the need of recalibration or scheduled maintenance is announced. Based on battery information and actual battery monitoring, system sends real time notification when solar powered stations require maintenance. Notification can be triggered by various combinations of measured data, metadata and even forecast data ingested into database from models.

Various reports can be generated from metadata in spreadsheet formats: actual station properties, history of station properties, observation and sensor information, stored meteorological elements description table etc.

Importance of the metadata management

Although at the first glance it may look like a lot of work with no additional information, it is very important to note that the metadata are essential in long-term data processing.

Below is example of series monthly averages of relative humidity in 6 years period. Starting from year 2011 there is changed character of data. Unfortunately, not many artificial changes in data are so clearly identifiable. Without proper and actualized metadata, it would be impossible to identify the change that occurred at beginning of year 2011. From actual metadata the user can identify that station location was changed to different near place and realize, that the sudden change was caused by this event.
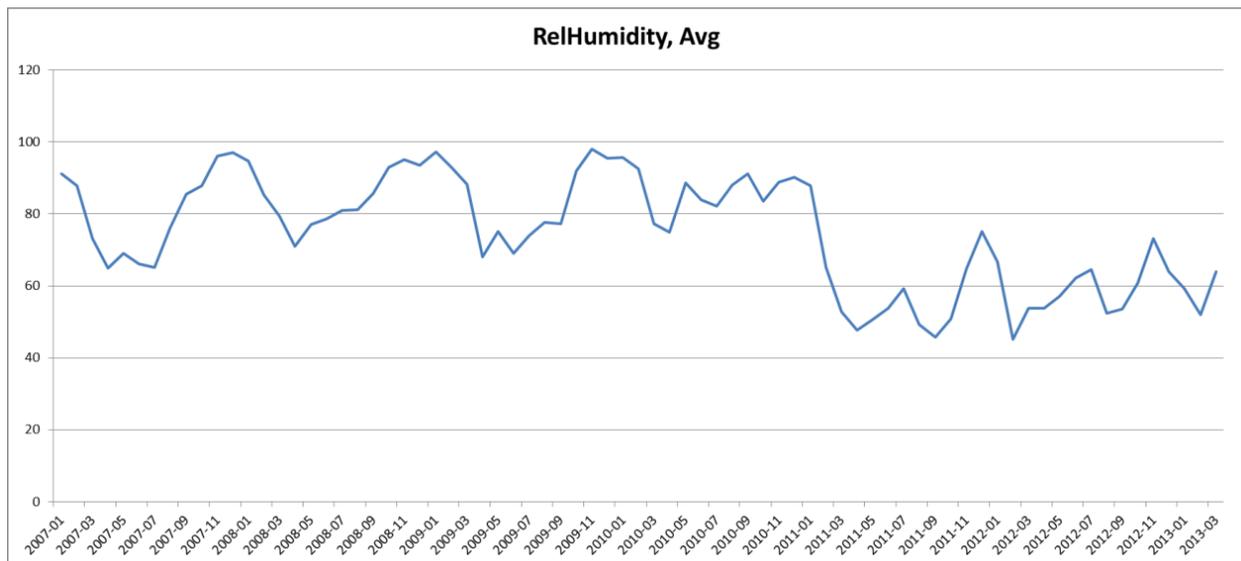
References

[1]   WMO (2001): Guide on Climatological Practices. WMO No.100. World Meteorological Organization, Geneva, Switzerland

[2]   WMO (2003): GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION. WMO/TD No. 1186. Coordinator of text: Enric Aguilar. World Meteorological Organization, Geneva, Switzerland