



The MeteoSwiss Data Warehouse System

As the national weather service, MeteoSwiss (Federal Office for Meteorology and Climatology) is charged with the collection, storage and administration of extensive meteorological and climatological data. To date, several billion data records have been archived, and measurements from numerous sources and in various formats arrive daily. With regard to the organization of efficient data processing, as well as a guarantee of rapid and flexible access for data analysis, the idea of a central platform with an integrated and consistent data pool becomes pivotal. MeteoSwiss has put this approach into practice by setting up and operating a data warehouse system. This embraces applications and a centralized and optimized database, the so-called data warehouse, for analytical and information purposes.

Project objectives

With the data warehouse project, MeteoSwiss has set up an efficient and extendable infrastructure that enables the collection and storage of meteorological and climatological measurements (basic data), as well as the relevant context information (metadata). It provides tools for the preparation and processing (aggregation, quality control, correction) of this data and forms the basis for made-to-measure data access for internal and external users. It is also possible, for the first time, to obtain up-to-the-minute data, as well as comprehensive time series covering decades, and to create products on the basis of all measurements and calculations.

Architecture of the MeteoSwiss data warehouse system

The MeteoSwiss data warehouse system corresponds to the architecture of a so-called corporation information factory, which combines data warehouse technology with elements of classic relational databases. It is subdivided into four levels (Figure 1): the source, staging, storage and access levels. A so-called metadata repository, which contains a context database that stores and administers the climatological metadata (e.g. measuring station information), is accessible for all levels. As a result, this guarantees that the same context data will be used at all levels.

The individual levels are briefly presented below:

Source level

The data flowing into the MeteoSwiss data warehouse comes from a number of different sources:

- observation systems (ground measuring networks, camera network, biometeorological measuring networks, radio soundings)
- numerical weather-forecasting models, amongst which the MeteoSwiss "Alpine model" (aLMo)
- the WMO Global Telecommunications System (GTS)
- weather radars and satellites
- data archives with historical data series

Staging level

The staging area of the data warehouse system contains the working database and all methods and techniques necessary for the collection, processing and loading, as well as the quality control and validation of the meteorological data.

Data storage level

The actual MeteoSwiss data warehouse is located at the data storage level. It consists of an analytical database that puts an integrated set of data at the disposal of meteorologists and climatologists, in order to assist them in weather forecasting and climate research. In addition, with the so-called data marts as selected sub-areas of the data warehouse, data pools, which are geared to functional demands and needs of certain users or fields, are created.

Data analysis level

The data analysis area contains the applications for carrying out queries on the data warehouse or the data marts. For example, with the CLIMAP application, measurements can be quickly and clearly visualized in the form of tables, charts or process diagrams. Access to the analytical database is not limited to internal users. Interested institutions and private individuals can also access the MeteoSwiss meteorological and climatological data via the Internet through the flexible and individually configurable CLIMAP-net software.

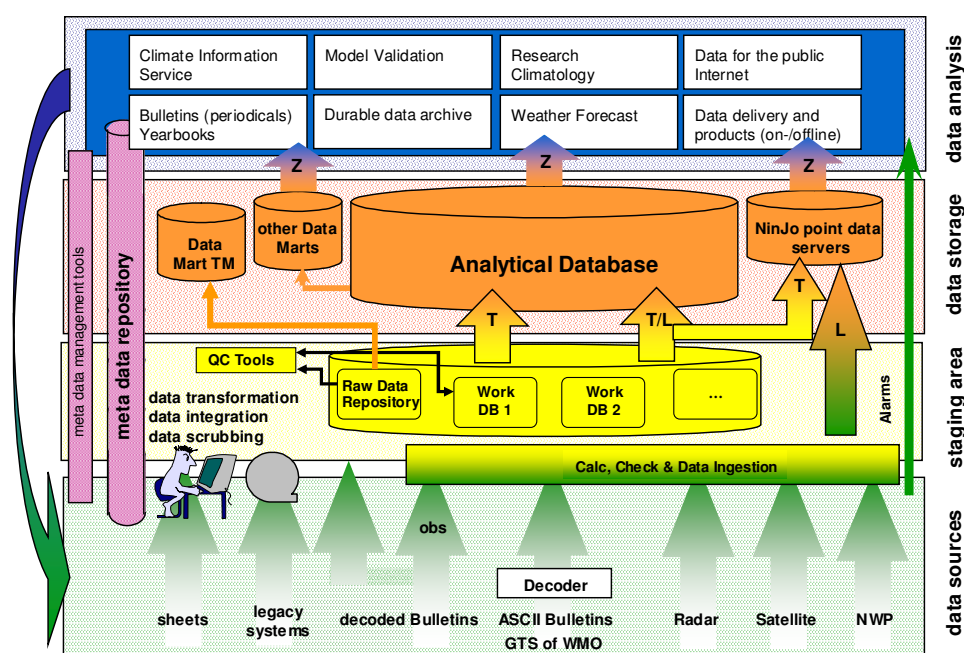


Figure 1: Concept architecture of the MeteoSwiss data warehouse system. The direction of the arrows shows the data flow (T - Transformation and loading, L - Loading without transformation, Z - ad hoc queries).

Project history

The MeteoSwiss data warehouse project officially started in 2000. The resulting reorganization of the database landscape was based on an expert's report by Professors Nievergelt and Zehnder (Department of Computer Science, ETH Zurich) together with an expert report by Professor Dittrich (Department of Informatics, University of Zurich). During the course of the development process, operational management of the new infrastructure was phased in gradually, so that various predecessor systems could be superseded. From the total of nine planned releases for the implementation of all the demands from the MeteoSwiss areas of activity, to date, six have been successfully carried out. The project is expected to be terminated towards the end of 2008.

Contact

MeteoSwiss, Christian Häberli, Krähbühlstrasse 58, CH-8044 Zürich