B/S Based Permanent Storage and Quality Control of Field Observation Data

Min Yufang, Zhang Yaonan, Long Yinping, Zhao Xueru and Aihong Zhang,
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences

According to the characteristics of field observation data and their storage, we must consider issues like data permanent preservation, quality control, data pre-processing and the speed of access. The paper puts forward a four-level data storage strategy and corresponding methods of data quality control. The four-level data storage includes a File Backup Database, an Original Database, a Processed Database and a Real-time Database. In order to guarantee the authenticity of data, the File Backup Database stored the original file as a backup in Data Center. After unifying data format, file data will be stored into the Original Database, and the database is mainly used as the contrast database of the Processed Database, which is helpful for tracing error data, inspecting quality control effect and evaluating the methods of quality control. The Processed Database stored data that have been done routine quality control and format conversion. The database is the most frequently used and provides historical data query, download and visual analysis. In order to improve the system response speed, the Real-time Database is used to store recent data and frequently used data. The four-level data management system also includes metadata database, log database, format conversion tools, quality control library and rights database as management library. Before a new type of data import into database, as long as data managers fill metadata into metadata database, design the table structure of data storage and set format conversion methods, quality control methods and visualization methods, the system will realize automatic storage of the real-time data, preliminary quality control and online visual display. Besides, the paper proposed a B/S based system to conduct observation data quality control, which includes two steps. Firstly, routine quality control including limit, range and consistency checking at a single station is conducted before data imported into the Processed Database, so that suspicious values or certain errors may have been identified. Then, meticulous quality control is carried out based on online visual data quality control and analysis. According to data request, the system will provide a variety of online quality control methods from quality control library, and will feed back interoperable diagrams to users, who can compare the effect of different quality control methods, and download the data of their interest. Take outlier detection as an example, the system can provide quality control methods including $3\delta'$ test method, Dixon' test method, and Grubbs' test method, etc. After quality control, the interoperable charts will be fed back to users, and this will help users to understand the overall of data and grasp the change rules over time and the intrinsic relationship between the data.

Keywords: Data storage, Quality Control, Visualization, Metadata, Field observation data