Development of SOA-Based Hydrological Data Compilation Management System

Yu Yang1, *, Xia Dazhong2, Jin Zheng2

1Haihe River Water Conservancy Commission, Tianjin, 300170, China
2College of Hydrology and Water Resources, Hohai University, Jiangsu, Nanjing, 210098, China
* e-mail: yuyang@hwcc.gov.cn

Abstract: As the results of hydrological data compilation are important fundamental information resources in China, how to efficiently manage them and provide data support for water science research is a major problem that needs to be addressed at present. In this paper, a hydrological database of Haihe River Basin was preliminarily established based on the compilation of the complete electronic data files, and a SOA-based management system of hydrological data compilation results was developed, featuring functions related to the compilation results. The research shows that the system boasts strong practicability and operability, and will be popularized in the later period, in order to improve the capacity and level of hydrological service, support water conservancy information and drive the modernization of hydrological service.

1. Introduction

Hydrological construction underpins the development of water conservancy and national economy in China. The integrated hydrological materials, the final product of hydrological data and one of the country’s fundamental information resources[1], have been widely used in flood control and drought relief, water conservancy project planning and design, water resources management and utilization, water environment protection, water science R&D and other national economic construction. They form an indispensably fundamental information resource. However, the constantly improving social and economic conditions and people’s living standard put forward increasingly high requirements for flood and drought disaster prevention, water environment monitoring and protection, soil and water conservation and ecological restoration, and water resources management, as well as hydrological data analysis, processing, storage and application services.

The vast majority of the compilation results of hydrological data over the years were generated in tables and then bound in a hydrologic year book. The hydrologic year book primarily contains accountability information and details in water level, runoff, sediment discharge, gradation of sediment grain, water temperature, ice run, water chemistry, groundwater, precipitation and evaporation, and corrections and additions of published data. It can be categorized into daily table, monthly and annual statistical table, abstraction of observed values, and extraction. Before 2000, the raw hydrological data were compiled into hydrologic year books, but were stored in a uniform spreadsheet format in the computer due to the subjective and objective conditions at that time; after 2000, hydrologic year books were published while hydrological data were kept, but there might exist inconsistent data formats for hydrologic year books; in recent years, hydrology construction has been integrated with modern
information technology, and the "Internet + Hydrology" pattern has been developed for the online compilation system whereby "Internet +" is applied into hydrological data compilation[2]. Therefore, how to digitize and manage the hydrological data before online compilation is an urgent problem.

2. Research objective
A SOA-based hydrological data integrated information service system was developed on hydrological data compilation results with complete electronic files generated after 2000, which would serve as a hydrological data compilation results management system shared by the Haihe River Water Conservancy Commission and its subordinate administrations. It prepared for the establishment of fundamental hydrological database system in Haihe River Basin, laid a foundation for development and application of hydrological information resources, provided information support for hydrological informationization and modernization, expanded the hydrological service field, improved hydrological service capacity and level, supported water conservancy informationization and promoted hydrological modernization services.

3. Overall architecture, design and development mode

3.1 SOA-based system architecture
SOA is a service-oriented enterprise application architecture and a distributed software architecture model. In this model, any business function is implemented as a service, and the different functions (services) of the application are linked through structures and contracts defined between these services. Thus, the application architecture can be viewed as an integration of services[3].

The SOA architecture consists of three participates -Service Provider, Service Consumer and Service Registry, and three basic operations[4]-Publish, Find, and Invoke. Service Provider publishes a service to a directory of Service Registry, and the Find operation is performed to search the service in the directory when Service Consumer calls the service. Service Registry returns information about the service, and Service Consumer calls the service according to the returned information. When Service Consumer gets information about the required service from Service Registry, communication is directly established between Service Consumer and Service Provider, not through Service Registry[5].

3.2 System design and development mode
The system was designed using the V2MC framework, which is evolved from MVC or Model View Controller, the abbreviation of Model - View - Controller. In this framework, business logic is separated from data and interface display and aggregated into one component. It allows not to rewrite the business logic while improving and customizing the interface and user interaction.

V2MC, as a new model, has a significant difference from MVC: in V2MC, the Controller is divided into DataController and ViewController, and the Model is categorized into DataModel and ViewModel. DataController is responsible for data interaction, and ViewController for business logic; DataModel is for original data to store, and ViewModel for view-needed data.

The system was developed using the mixed mode of C/S and B/S, where the latter is main mode, consisting of system application layer, system service layer and system data layer.

- System application layer: B/S mode is adopted to develop and integrate the management system for the query of various reports and hydrological elements; C/S mode is employed to develop the importing function of all hydrological data.
- System service layer: Based on an integrated supporting platform, it mainly provides system application-level services through three components: core components, public service components and professional service components. Core components form the necessary environment and services to support platform operation, including core service components, platform runtime components and other key components. Public service components include Web service components, GIS service publishing components, application management
components, user management components, resource management components and database management components. Professional service components comprise hydrological report components and hydrological data chart components.

- System data layer: Several databases are designed according to the requirements of the platform, including system configuration database, historical data compilation database and real-time hydrological information database.

### 4. Main functions of the system

The functions of the system can be categorized in data import, data cleaning, data maintenance, data analysis and data reporting by business requirements.

The basic database tables of the system are in accordance with *The Standard for Structure and Identifier in Fundamental Hydrological Database* (SL324-2005)[6] (consisting of basic information tables, abstraction of observed values, daily tables, ten-day tables, monthly tables, annual tables and measurement tables) and *The Structure and Identifier for Real-time Hydrological Information Database* (SL323-2011)[7] (mainly including precipitation table, daily evaporation table, river regime table, reservoir regime table, weir gate regime table, tidal regime table), and the table structure was expanded as per the application requirements of the system to construct the basic database table suitable for the system.

![Main functions of hydrological data compilation results management system](image)

#### 4.1 Data import

By using C/S mode, data import is developed to put the typesetting files or the Excel data of a hydrological yearbook into database. The data of the hydrological yearbook include accountability information and details in water level, runoff, sediment discharge, water surface evaporation, precipitation, ice run and gradation of sediment grain[8]. If the typesetting files of the year is opted, the system will automatically analyze and store the files. The data import process was shown in Figure 2.
4.2 Data cleaning

The typesetting files of the hydrological yearbook are imported and automatically analyzed, and then undergo "data cleaning" after storage in the database. According to the characteristics of hydrological data, the following data cleaning process was implemented to clean the data.

- **Data integrity**: Check for missing data in a database. Integrity check is mainly for year, station, table type, and data integrity in a table (via checking whether there is any vacancy in the data in the table).

- **Data rationality**: whether the logical relationship between similar data in the database is correct. Rationality check comprises eigenvalue rationality check and single station rationality check. The former is mainly to logically check similar statistical values, such as monthly statistical values which should be less than or equal to annual statistical values; the latter is to check the logical relationship between similar hydrological elements, e.g. whether the daily average water level comes between the lowest water level and the highest water level in the month.

- **Data consistency**: whether the logical relationship between different types of data in the database is correct. Under normal circumstances, the daily average water level of a single station corresponds to the daily average runoff.

- **Data relevance**: whether the logical relationship between data of the same type in stations of the database is correct. If a river section has upstream and downstream sub-sections, the peak runoff of the downstream sub-section should be smaller than that of the upstream sub-section in case of no tributaries or rainfall.

4.3 Data maintenance

After the electronic data of the hydrological yearbook is stored in the database, there may be problems in analyzing the data using software, so it is necessary to re-check the data after storage to ensure its accuracy.

4.4 Data analysis

Data analysis was composed of comparison of single element data from multiple upstream and downstream sub-sections, frequency analysis on rainfall and runoff, calculation of statistical parameters and design values; calculation of the maximum, minimum and average eigenvalues, design of flood process lines by using statistical parameters and typical calculation, regression statistical analysis on multi - station data, and output regression equation.
4.5 Data report

The system provides queries of complex conditions comprising \{single (multiple) stations×single (multiple) time frames×single (multiple) elements\}, and the mixed mode of structured and unstructured data. The query results are displayed in graphs and tables and the query result structure is optional.

It can be combined with the category query to support query results export, where the results of specific queries can be exported in a fixed format or directly print in a fixed format report.

5. Conclusion

The results of hydrological data compilation are one of the fundamental information resources for a country. In this paper, the compilation results of complete electronic data files were utilized to develop an SOA-based hydrological data compilation management system for daily business needs, including data import, data cleaning, data maintenance, data analysis and data reporting of yearbook electronic files. At this point, the fundamental hydrological database of Haihe River Basin was preliminarily established. But how to convert the compilation results of paper hydrological data into the online database data as backup is an urgent problem to be solved. Currently, the wide application of scanners and the improvement of computer and Optical Character Recognition (OCR) technology make it possible to quickly and accurately convert paper forms into editable electronic data, which will be investigated emphatically in the subsequent phase. Therefore, a long series and all-factor fundamental hydrological database was constructed, providing information support for hydrological informationization and modernization, expanding the hydrological service field, improving hydrological service ability and level, supporting water conservancy informationization and driving hydrological modernization service.

References

[1] Zhang S.A., Wu L.u., Lin W., et al. (2006) Development of the Technology of Hydrologic Data Processing and Database in China. Hydrology, 26 (3): 48-52.
[2] Pan Z.B., Ke B.L., Chen G.Q., et al. (2018) Application of Partial Duration Series and Copulas in Deriving Design Flood of Reservoir. Hydrology, 38 (2): 81-84.
[3] Jian B., Zuo R.G., Yan G.R., et al. (2007) SOA-based Application Integration System of Small and Medium-sized Manufacture Enterprises. Computer Engineering, (05):243-245.
[4] Holger S. (2005) Modeling and Model Checking Web Services. Electronic Notes in Theoretical Computer Science, 126.
[5] Chi W.X. (2008) Research on Service-Oriented Architecture Building Development Technology. Application Research of Computers, 25(6):1717-1719.
[6] Ministry of Water Resources of the People's Republic of China. (2005) The Standard for Structure and Identifier in Fundamental Hydrological Database: SL 324-2005. Beijing: China Water & Power Press.
[7] Ministry of Water Resources of the People's Republic of China. (2011) Structure and Identifier for Real-time Hydrological Information Database: SL 323-2011. Beijing: China Water & Power Press.
[8] Yangtze River Water Conservancy Commission of the Ministry of Water Resources. (2012) Code for Hydrologic Data Processing: SL 247-2012. Beijing: China Water & Power Press.