Integrated Technical Pathways for Multi-Source Heterogeneous Water Conservancy Data: A Case Study of the Ili Prefecture Business Data Access Scheme

Chenchen Cui *, Jing Bai , and Xiaoheng Chen

Information Center, Yellow River Conservancy Commission, Zhengzhou, China;

Abstract. Water conservancy data — serving as the core foundation for water resources management, flood and drought disaster prevention, ecological protection, and hydraulic engineering operations—originate from diverse sources, vary significantly in structure, and exhibit different requirements for timeliness. These characteristics result in typical multi-source heterogeneity. Achieving effective integration of such data has become a key technical bottleneck in building smart water conservancy systems. Using the 2024 water conservancy informatization data aggregation scheme of the Ili Prefecture in Xinjiang as an empirical case, this study systematically analyzes the region' s data integration practices across eight major business domains, including river and lake supervision, project construction management, irrigation district management, and rural drinking water safety [2]. Three representative data access modes are summarized: real-time access of monitoring data, exchange-based integration of business system data, and review-based consolidation of historical data. A quality control mechanism centered on "reporting rate statistics" and an abnormal-data governance workflow are refined. On this basis, a three-stage implementation framework — " specification formulation, phased access, and dynamic quality inspection" —is proposed. This framework provides a replicable and scalable technical pathway for similar regions and offers methodological guidance for promoting water-related data capitalization, business collaboration, and intelligent decision-making.

Keywords: Multi-source heterogeneous data; data integration; water conservancy informatization; Ili Prefecture; data quality control.

1. Introduction

With the deep integration of new-generation information technologies—such as the Internet of Things, cloud computing, and big data—into the water conservancy sector, the scale, diversity, and value of water-related data have increased sharply [1][9]. However, such data inherently possess three major characteristics: multi-sourced (originating from hydrological stations, meteorological stations, engineering sensors, business systems, and manual inputs), heterogeneous (including real-time streams, geospatial data, relational business data, and textual reports), and decentralized (stored across departments, administrative levels, and independent systems). These characteristics lead to challenges such as inconsistent technical protocols, lack of unified standards, uneven data quality, and barriers to data sharing and circulation [6][10].

The Ili Kazakh Autonomous Prefecture in Xinjiang is rich in water resources, has a comprehensive water conservancy business system, and has urgent needs for informatization development. Its 2024 data aggregation advancement plan systematically designed access tasks for eight categories of core business data, including river and lake supervision, project construction management, flood and drought disaster prevention, irrigation district management, rural drinking water safety, groundwater monitoring, reservoir safety operation, and underground well management. This provides a valuable full-scenario practical case for studying multi-source heterogeneous water-related data integration. This paper aims to deconstruct the technical realization pathways of the scheme, extract its integration models, quality control methods, and implementation strategies, and offer systematic and actionable references for data governance in grassroots water conservancy departments.

2. Multi-Source Heterogeneous Data Integration Models: The Practice of Ili Prefecture

Based on differences in data sources, forms, and business requirements, the Ili Prefecture scheme adopts differentiated integration modes, which can be summarized into the following three categories [7].

2.1 Real-Time Access of Monitoring Data: Example of River and Lake Supervision

Monitoring data mainly refer to time-series measurements automatically collected by sensors and generated in near real time, such as water level, discharge, rainfall, water quality, and video images [4][5]. These data have extremely high requirements for timeliness, continuity, and reliability.

Protocol alignment and specification first: for 18 key monitoring cross-sections of important rivers, the project first developed the Ecological Base Flow Monitoring Data Access Specification. It defines data formats, communication protocols (MQTT, HTTP), transmission frequency, measurement units, etc., providing a unified technical benchmark for integrating heterogeneous monitoring equipment from various manufacturers and construction periods.

Centralized platform-based receiving: a unified data receiving platform or IoT platform is established. Monitoring station data are transmitted to the platform via 4G/5G or BeiDou satellite communication in accordance with the defined specifications [4]. The “collection and receiving platform” mentioned in the scheme performs this core function.

Real-time processing and storage: the platform parses and decodes raw data, performs basic validation (e.g., range checks), and writes valid results into a time-series database or a specialized water conservancy database for query, visualization, and early-warning applications.

Overlapping-data identification: for hydrological stations that may overlap with those managed by the regional hydrology bureau, the scheme requires clarification of source relationships to avoid duplicate storage. This process may involve assigning unique source identifiers or implementing primary-secondary synchronization mechanisms.

The technical focus is on establishing unified specifications and a highly available receiving platform that can handle complex field environments and high-concurrency transmission.

2.2 Exchange-Based Integration of Business System Data: Example of Project Construction Management

Business system data mainly refer to structured data stored in information systems at various levels of water conservancy authorities, such as project information, approval workflows, and documentation. Their integration emphasizes completeness, consistency, and standardization.

Table-structure alignment and mapping: the access of project construction management data explicitly follows the database table structure of the autonomous-region-level system. This upward-alignment approach ensures consistency with upper-level standards and removes structural barriers to future reporting and data sharing.

Data exchange and synchronization: using ETL tools or customized data-exchange services, data pertaining to Ili Prefecture are regularly (or event-driven) extracted from the regional system. Requirements for database connection parameters and data dictionaries ensure secure and accurate data exchange.

Local database construction and enhancement: after receiving upper-level data, a local database with identical structures is built. Additional information required for local management is supplemented to form a complete dataset that is both standard-compliant and locally applicable.

The core lies in the unification of data standards (structural alignment) and secure, controllable data-exchange mechanisms.

2.3 Review-Based Consolidation of Historical Data: Example of Flood and Drought Disaster Prevention Station Ledger

Historical and legacy data often suffer from incompleteness, errors, and inconsistent standards, requiring substantial cleaning, verification, and consolidation.

Ledger creation and checklist management: the scheme requires reviewing information on mountain flood monitoring stations, hydrological stations, and meteorological stations to compile a comprehensive station-network ledger—essentially clarifying “what data sources exist, where they are, and their status.”

Multi-source comparison and manual verification: station lists from different systems (water conservancy, meteorology) are compared to identify duplicate, missing, or inconsistent entries. Business personnel verify uncertain cases.

Data completion and relationship building: based on the verified ledger, individual monitoring data are accessed or supplemented. Spatial location, management attributes, and monitoring indicators are associated to form a coherent knowledge system of the monitoring network.

The key lies in a business-driven verification process and checklist-based governance approach.

3. Collaborative Quality Control in Data Integration

Data integration is never a one-time task. Continuous quality control is essential for data usability. The Ili Prefecture scheme demonstrates a quality-control approach emphasizing “quantified monitoring, visualized issues, and collaborative rectification.”

3.1 Reporting-Rate Statistics Dashboard: Enabling Visual Quality Monitoring

For each data category (river and lake sections, irrigation measurement points, groundwater monitoring stations, rural water plants, etc.), dedicated dashboards for reporting rate and compliance statistics are required.

Fig.1 dynamically display real-time status, historical trends, and rankings of reporting compliance across data sources and counties. They transform previously invisible integration processes and quality conditions into intuitive, measurable visual charts, enabling managers to quickly detect anomalies (e.g., sudden drops in reporting rate) and identify responsible units (contact persons are labeled on the interface) [3]. BI tools compute reporting indicators from databases and message queues in real time, while frontend chart libraries render visualizations—providing a lightweight yet highly effective management tool.

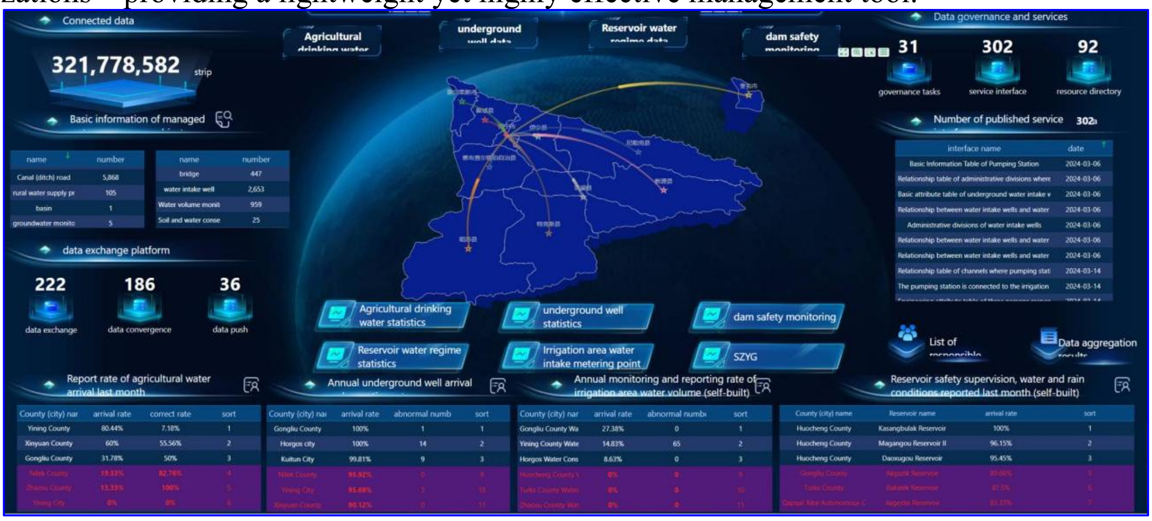

Fig.1 Reporting-Rate Statistics Dashboard

3.2 Abnormal-Data Detection and Collaborative Rectification: Example of Underground-Well Data

Typical quality issues include low reporting rates for real-time monitoring (shown as in Fig.2), numerous abnormal values, and repetitive or missing daily data, which distort monthly and annual statistics.

Through reporting-rate monitoring, “low reporting rate” issues are identified; rule-based validation detects “abnormal data”; comparative logic uncovers “duplicate or missing pushes.” The scheme further clarifies station-ledger discrepancies through business-driven verification and assigns rectification responsibilities (“county-level verification through water resources offices”). Technical support is provided to counties, and data-production logic is optimized (e.g., adding monthly and annual statistical push functions) to prevent discrepancies at the source.

This forms a closed loop: technical detection → business confirmation → management assignment → technical support for rectification.

Statistical page for groundwater data arrival rate in 2025 please select 2025

County (city) name	2025			
	Number of registrations (eyes)	Average number of reports received	Report arrival rate (%)	abnormal information
Yining City	116	111	95.69%	3
Kuitun City	525	524	99.81%	9
Horgos city	188	188	100%	14
Yining County	237	232	97.89%	0
Qapqal Xibe Autonomous C	403	402	99.75%	1
Xinyuan County	81	73	90.12%	0
Gongliu County	50	50	100%	1
Zhaosu County	79	76	96.20%	0
Nilek County	98	94	95.92%	0
Huocheng County	651	642	98.62%	0
Turks County	47	46	97.87%	0
total	2,475	2,438	98.51%	

Fig.2 Reporting Rates for Real-time Monitoring

4. A Three-Stage Technical Implementation Framework: Specification – Access – Dynamic Quality Inspection

Synthesizing Ili Prefecture’s practices, a spiral three-stage implementation framework (Fig.3) suitable for multi-source heterogeneous

water conservancy data integration is extracted.

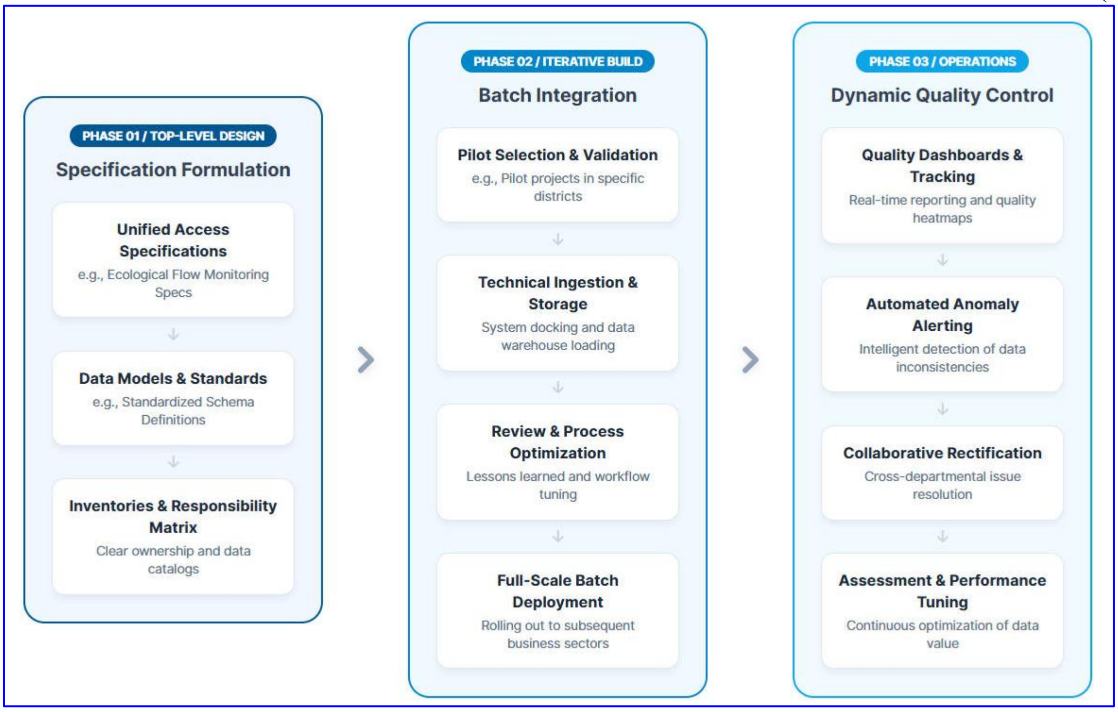

Fig.3 Spiral Three-stage Implementation Framework

4.1 Stage One: Specification Formulation (Top-Level Design)

Key tasks include establishing the “rules of the road” for integration: developing technical specifications for access, defining unified data models and coding standards (e.g., reservoir codes), and clarifying business ledgers.

Ili practices include developing the Ecological Base Flow Monitoring Data Access Specification, designing local databases following upper-level table structures, and verifying station ledgers for flood and drought disaster prevention.

4.2 Stage Two: Phased Access (Iterative Implementation)

Based on the principle of “from easy to difficult, from pilot to full coverage,” data access is implemented by business category and region.

Examples: Irrigation-district data began with a pilot in Gongliu County, followed by expansion to Yining County and eventually the entire prefecture. Rural drinking water data were integrated in batches.

4.3 Stage Three: Dynamic Quality Inspection (Continuous Operation)

After access, continuous monitoring and governance mechanisms are required. Visualization tools facilitate quality monitoring, while cross-department workflows enable issue identification, feedback, and rectification.

Quality dashboards, notifications, and complete workflows—from ledger review to statistical logic optimization—represent Ili’s iterative governance practice.

These three stages are not strictly linear but cyclic and mutually reinforcing. For example, issues identified in dynamic inspection may require revisiting and revising the initial specifications.

5. Conclusion and Future Outlook

The Ili Prefecture data aggregation scheme provides a practical and systematic regional-level example addressing the challenges of integrating multi-source heterogeneous water-related data. Its key insights include:

Model-based handling of heterogeneity: different data types—real-time monitoring, business-system data, and historical ledgers—require differentiated integration modes to ensure efficiency and effectiveness.

Visualization-driven quality control: transforming reporting rates and quality indicators into management dashboards enables proactive rather than reactive data governance.

Framework-based implementation assurance: the three-stage framework of "specification formulation, phased access, and dynamic quality inspection" organically connects standards, implementation, and operations, providing methodological support for complex integration projects [8].

Looking ahead, water-data integration technologies will evolve toward greater intelligence and automation. AI-based abnormal-data detection and repair, knowledge-graph-driven semantic association and fusion, and blockchain-based trustworthy data exchange mechanisms will be key directions for next-generation platforms [2]. The Ili experience provides a solid data foundation for these advancements, and its emphasis on "business-technology collaboration" and "continuous governance" will remain fundamental principles for future smart-water-conservancy data systems.

References

[1] He, X.; Zhang, M. Overview of Research Progress on Water Conservancy Big Data in China. In: Proceedings of the 2024 4th International Conference on Public Management and Big Data Analysis (PMBDA 2024); Atlantis Press, 2025.

[2] Li, Q.; Ma, Z.; Li, J.; Li, W.; Li, Y.; Yang, J. Overview of the Research Status of Intelligent Water Conservancy Technology System. Applied Sciences, 2024, 14(17): 7809. doi:10.3390/app14177809.

[3] Jamal, E.; Moumen, A.; Rissouni, Y.; Chao, J.; Tahi, A. Big Data Analytic and IoT for Water Resources. In: Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021). SCITEPRESS, 2022, pp. 433 – 439. doi:10.5220/0010736000003101.

[4] Ansari, S.A.; Vidyarthi, V.K. Use of Internet of Things in Water Resources Applications: Challenges and Future Directions—A Critical Review. Discover Internet of Things, 2025, 5(1): 96. doi:10.1007/s43926-025-00193-7.

[5] Zhang, S.; Ting, L.; Chao, W. Multi-Source Data Fusion Method for Structural Safety Assessment of Water Diversion Structures. Journal of Hydroinformatics, 2021, 23(2): 249 – 269.

[6] Shen, J.; Zhou, J.; Herman, L.; Řezník, T. Constructing the CityGML ADE for the Multi-Source Data Integration of Urban Flooding. ISPRS International Journal of Geo-Information, 2020, 9(6): 359.

[7] Zhang, Z.; Wang, S.; Xia, J.; et al. Multi-Source Data Fusion and Hydrodynamics for Urban Waterlogging Risk Identification. International Journal of Environmental Research and Public Health, 2023, 20(3): 2528.

[8] Li, F.; Xie, L.; Huang, W. Practice and Key Technologies of Integration and Sharing of Water Resources and Data Resources Based on Cloud Computing. Inventi Impact: Cloud Computing, 2022, July – September issue.

[9] Chen J., Deng M., Wang H. A Review of Water Conservancy Big Data Research[J]. Advances in Water Science, 2017, 28(4): 622 – 631.

[10] Qin H., Zhang J., Gao J., et al. Data Sharing Technology Among Heterogeneous Hydrological Information Systems[J]. China Flood & Drought Management, 2018, 28(11): 40 – 43, 48.