



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Design and Application of an Integrated System for Big Data Management in University Auditing based on Data Middle Platform

Linjun Kong

Office of Digitalization, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang, China

ABSTRACT: The integrated management system for audit big data is an unified management platform established based on the successful practice of the "data hub" in a university. This system adheres to the principles of "comprehensiveness, accuracy, and intelligence." Through a new-generation, all-encompassing data center that has undergone data governance, it comprehensively integrates various underlying data types. It utilizes reporting tools to visually present and verify all types of data or information required for submission, ultimately facilitating the online review and submission of reported data. Additionally, it archives and preserves the submitted data over the years, enabling convenient subsequent access, utilization, and statistical analysis of historical data.

KEYWORDS: Audit big data, Data hub, Data submission, System.

I. BACKGROUND

In the current information era, big data, akin to other natural resources, has emerged as a crucial strategic asset. It views the essence of transactions as a collection of data, enabling the exploration and processing of relationships within vast amounts of data through data-driven thinking and advanced information technology. As research progresses, big data technology is playing an increasingly significant role in various fields, including social production. For audit informatization, the development of big data technology has revolutionized data auditing, leading to a series of transformations in audit objectives, content, analytical techniques, and audit thinking patterns. It represents a novel electronic audit approach for massive data, requiring the handling of both structured and unstructured data, thereby posing higher demands on all aspects of auditing. Leveraging big data analytics technology to build a big data audit verification system shifts audit methodologies from confirmatory analysis to exploratory analysis. This transformation of audit models, from traditional paper-based accounting ledgers to big data auditing, facilitates data forecasting by audit departments.

On February 18, 2021, the Zhejiang Provincial Party Committee and Provincial Government emphasized at the Provincial Digital Reform Conference [1] that "Zhejiang's digital reform aims to adapt to the digital wave, drive the gradual infiltration, widespread adoption, and full integration of information technology, and ultimately establish digital awareness and thinking, cultivate digital capabilities and methodologies, and construct a digital governance system and mechanism, proactively leading the global digital transformation and unlocking new spaces for value creation." The conference highlighted the need to "maximize the openness of data assets, promote correlated data applications, and stimulate the amplifying, overlapping, and multiplying effects of data as a production factor on the economy and society, empowering both the reform process itself and society, enhancing governance capabilities." Accelerating the construction of an integrated and intelligent public data platform, adhering to the principles of "unified planning, support, architecture, platform, standards, construction, management, and operation," employs an integrated data catalog and leverages public application support components for intelligent analysis, evaluation, and research of data and information, driving scientific decision-making and efficient execution. At the top level of data, it aggregates public basic data, production factor data, scientific and technological innovation data, and comprehensively promotes

the digitization of social undertakings. Focusing on 12 areas such as education and employment, it expedites the formulation of system integration lists and data sharing lists, advancing innovations in public service delivery.

In September 2020, the Zhejiang Provincial Department of Education issued a notice requiring 70 public universities across the province to establish a regularized mechanism for submitting audit data, periodically providing relevant electronic data. However, the decentralization of audit business data across various departments within universities makes data collection and management time-consuming. Additionally, the diversity of data types, including system data, spreadsheets, text documents, and others, coupled with low levels of standardization and normalization, poses significant challenges for the reporting of audit data.

II. CURRENT RESEARCH STATUS

In 2015, the State Council of China issued the "Action Outline for Promoting the Development of Big Data," advocating for "comprehensively advancing the development of big data in China and implementing the national big data strategy." The "Opinions of the State Council on Strengthening Audit Work" also emphasized the need to "explore ways to apply big data technology in audit practices, intensify the comprehensive utilization of data, and enhance the capabilities of using information technology to investigate issues, conduct evaluations and judgments, and conduct macroeconomic analyses."

Scholar Earley noted that "in the new era, big data has transformed the way of auditing, presenting both challenges and opportunities" [2]. BrownLiburd suggested that "to successfully implement data auditing, it is necessary to develop tools or software platforms, train practitioners' professional skills, and support the rapid development of audit data" [3]. Gepp A and other scholars have also delved into the connection between auditing and data processing systems, discussing issues related to audit methods and content[4]. Domestic researcher Qin Rongsheng elaborated on the impact of cloud computing and big data technologies on audit methods, audit sampling techniques, audit report models, and other techniques and methodologies[5]. Xu Chao et al. summarized, compared, and analyzed research related to big data auditing in data collection, storage, analysis, and visualization[6]. Liu Xing, Niu Yanfang, and Tang Zhihao analyzed the data, technical, and application characteristics of big data auditing, engaging in a fruitful discussion of several key issues facing big data auditing[7]. Zheng Wei et al. assessed the feasibility of data-driven audit models in a big data environment, proposing that a multi-dimensional analysis system could be built upon big data audit platforms to support horizontal and vertical applications[8]. Yao Junke et al. introduced a university internal audit framework based on big data, using the internal audit of scientific research funding in a university as an example to establish an overall analysis-internal analysis model for audit data[9].

In summary, while academia has conducted extensive research on big data auditing in the field of public management, relatively few studies have focused on big data auditing in the educational sector. Furthermore, most of the existing research is theoretical in nature, with limited exploration into how to establish a complete link for big data auditing and verification in universities, build a big data audit verification system, and implement it in practice. This gap in research highlights the significance of the present study.

At the end of 2020, a university in Zhejiang, leveraging the achievements of its earlier "Data Hub" - the comprehensive data center, explored the construction of an integrated application scenario for big data management in university audits, supported by the university's full range of data. It established a normalized audit data submission mechanism at both the provincial and university levels, realizing a full-link closed-loop management process of "data collection - verification - standardization - storage - submission." This addresses the traditional, human-centric issues in audit data submission, effectively enhancing the quality and efficiency of audit informatization work, and innovating the mechanism and mode of normalized audit data submission.

III. DESIGN OF AN INTEGRATED SYSTEM FOR AUDIT DATA MANAGEMENT IN UNIVERSITIES: ZHEJIANG CASE STUDY

1. System design approach

In response to the audit data submission requirements from the Zhejiang Provincial Department of Education, targeting 20 business domains within the university (including academic affairs, assets, library, personnel, scientific research, student affairs, finance, procurement, etc.), the integrated audit data management system must rely on the existing "Data Hub" construction achievements of the university to achieve unified acquisition, standardized storage, and secure storage of audit data. Additionally, it establishes a normalized audit data submission mechanism at both the provincial and university levels, realizing a full-link closed-loop management of audit data from its business source acquisition to submission and analysis. The overall design approach for the project is as follows:

1.1 Incremental development and iterative upgrades based on the university's existing "Data Hub"

The construction of the integrated audit data management system is designed to incrementally develop and iteratively upgrade based on the university's existing new-generation comprehensive data center and data governance achievements, specifically addressing audit data submission requirements. This approach allows for the reuse of previous data-level achievements, enabling horizontal expansion of the existing data architecture. It significantly reduces the overall development workload while ensuring the continuous operation of the current data system.

1.2 Achieving unified data acquisition, centralized management, and standardized output for various data sources

The audit data submission covers data from 20 business domains of the university, with some data residing in business management systems and others unsupported by such systems, existing offline in Excel files. In this scenario, different acquisition architectures are adopted for different data sources to achieve unified acquisition: business data is uniformly acquired and standardized into the database through the "Data Hub", while offline data is centrally imported through a reporting system. The system performs centralized management, including data standard management, data model management, etc. on the uniformly acquired data, builds an audit data standard warehouse and audit-specific data marts, interfaces with the provincial education department's data portals, and provides standardized data output.

1.3 Constructing data verification models for full-process, visual monitoring of data quality

Before audit data submission, the quality of relevant data must be verified. Data quality issues can be divided into structural and content-related issues. For structural data quality issues, after unified data storage, relevant software tools in the "Data Hub" perform structured cleaning. For content-related data quality issues, data to be verified is centrally presented in the form of reports, and the relationships between data tables are validated. A comprehensive data verification model is established, which alerts for abnormal data, allowing relevant reviewers to browse, review, modify, and confirm the data content. This complete, visual closed-loop process for data quality monitoring ultimately generates a high-quality audit data topic library and completes data submission.

1.4 Conducting multi-dimensional visual analysis of audit data through reporting tools to guide relevant school decisions

Apart from managing the full-link closed-loop of "acquisition - verification - standardization - storage - submission" of audit data, the integrated audit data management system also utilizes reporting tools to conduct multi-dimensional, in-depth, and visual analysis of annual audit data. This enhances the comprehensive application of audit data in correlation analysis, trend prediction, and other aspects, effectively promoting the digital transformation of audit operations and improving the precision management and scientific decision-making capabilities of the university's audit department.

2. Design of system technical architecture

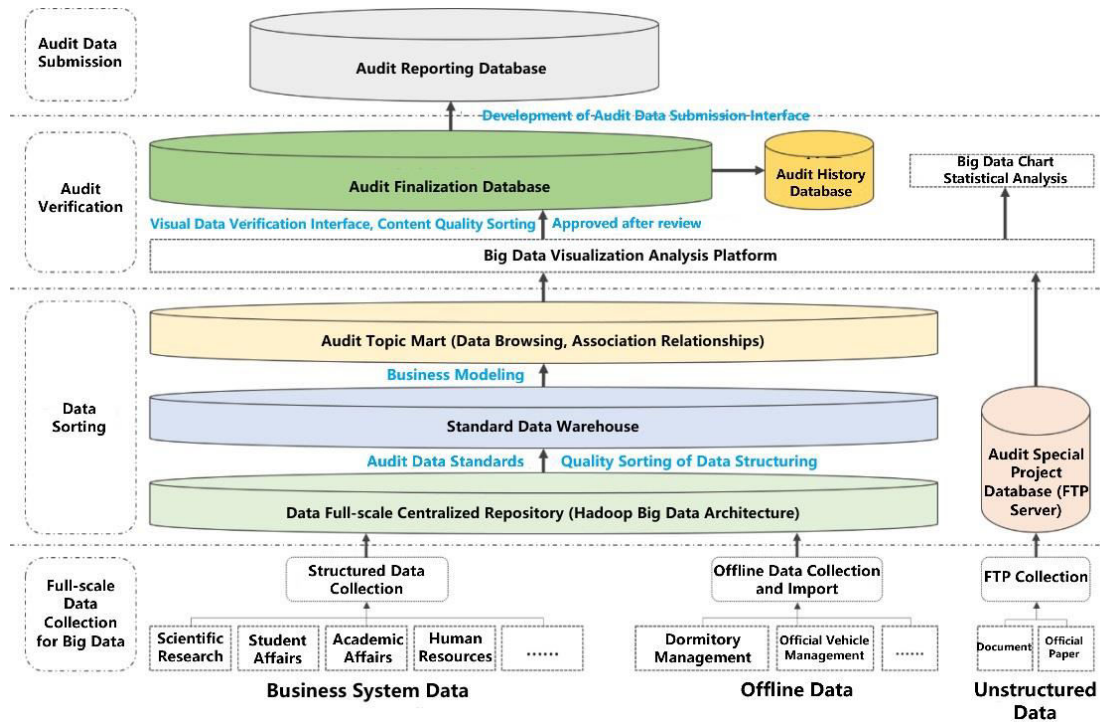


Figure 1: The Architecture Diagram of the University's Big Data Audit Verification Platform

The figure above depicts the architecture of a university's big data audit verification system. As shown, the data audit verification system built upon big data computing technology can be divided into four main layers based on the data architecture hierarchy: Acquisition Layer, Consolidation Layer, Audit Verification Layer, and Audit Reporting Layer.

2.1 Acquisition layer

Primarily focused on gathering various types of data from various business system data sources within the university that are required for auditing, offline spreadsheet data, and unstructured data such as files and texts. This layer performs unified and comprehensive data acquisition, employing different acquisition architectures tailored to different data source types. Additionally, considering the regular audit data submission requirements, the acquisition mode can be configured as either full or incremental.

2.2 Consolidation layer

Initially, business data from the Acquisition Layer is centralized 1:1 into a comprehensive data repository based on Hadoop's big data storage architecture, supporting distributed deployment and accommodating multiple data source types. Subsequently, this layer standardizes audit data, enhances structural quality, and integrates data structures. It serves as the primary level for managing audit data as assets. After integrating data according to audit data submission requirements, it is stored uniformly to facilitate subsequent querying and analysis. The standard data warehouse, as the core of centralized data storage and integration, facilitates data sharing and eliminates issues of redundant storage and processing across systems. The audit topic mart mainly stores raw and derived data sourced from the standard data warehouse, processes it according to the business needs of audit data submission, and provides the required result data.

2.3 Audit verification layer

Leveraging a big data visualization analysis platform, this layer develops visual verification interfaces for result data within the audit topic mart, facilitating auditors in reviewing data content for audit verification. If content-related data quality issues are identified, feedback can be promptly provided to the data source department for correction. Upon successful verification, the data flows to the audit finalization database, where its structure aligns perfectly with the audit data submission requirements. Simultaneously, a historical data backup for the current year is created and stored in the audit history database, facilitating historical data retention and future reference. For statistical analysis of audit data, this platform enables deep analysis of consolidated data, presenting processing results in various forms, including fixed reports, multidimensional analysis, and flexible queries.

2.4 Audit reporting layer

Tailored to the requirements of audit data submission, this layer custom-develops lightweight API interfaces, text uploads, direct database connections, and other methods to facilitate data reporting operations.

3 System function design

3.1 Overall functional module diagram

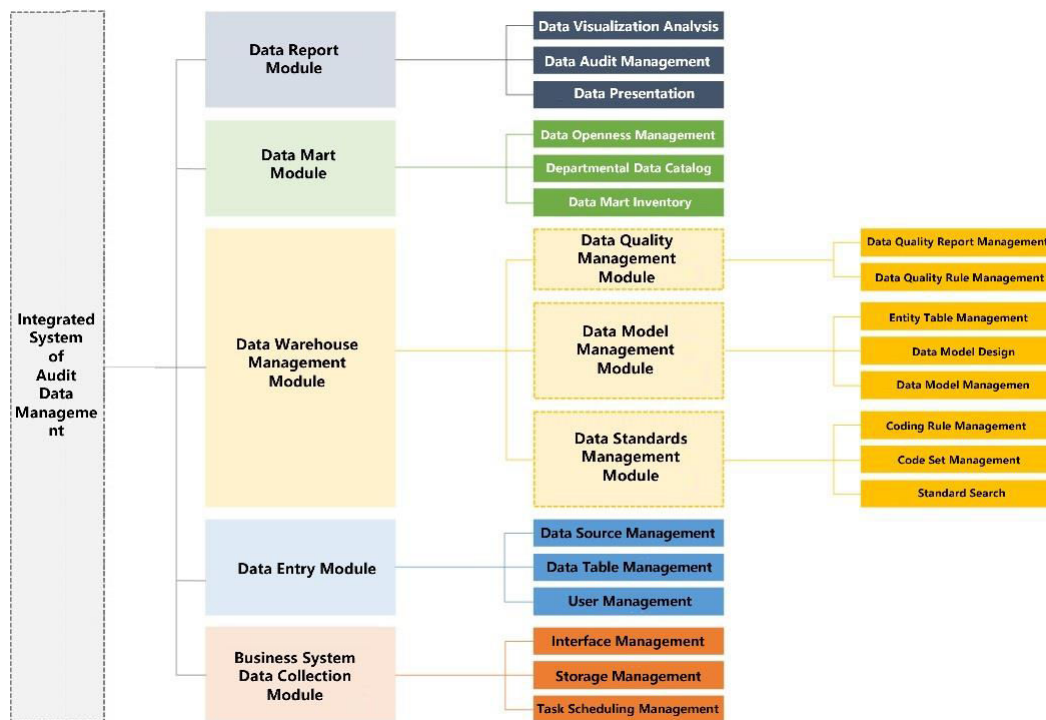


Figure 2. Overall Functional Module Diagram of the System

3.2 Detailed system function description

(1) Business System Data Collection Module

This module primarily focuses on the unified collection and aggregation of data from various business management systems existing within the school. It encompasses functions such as interface management, storage management, and task scheduling management.

1) Interface Management. Interface management involves managing both custom interfaces for data acquisition from business systems and developer interfaces. Firstly, the custom interface module is utilized for managing

created ETL (Extract, Transform, Load) interfaces. It displays the current status of established ETL interfaces and supports searching ETL interfaces based on interface name/description, interface type, and interface status. Secondly, the developer interface module manages interfaces created by developers. It presents information on all interfaces created by developers and allows searching based on developer names.

2) Storage Management. Storage management supports the management of the storage process for already connected data, enabling searches based on the storage procedure name, description, and database type.

3) Task Scheduling Management. Task scheduling management oversees the current data acquisition scheduling tasks, allowing for the creation, categorization, and deletion of acquisition interface tasks. It supports searching by scheduling name, associated interface name, scheduling type, and scheduling status. Additionally, it enables detailed viewing of interface task runtime status through logs.

(2) Data Reporting Module

This module is responsible for the unified reporting, importing, and management of offline Excel data that lacks support from business management systems. It includes functions such as data source management, data table management, and user management.

1) Data Source Management. Manages data sources for the data reporting module, currently supporting structured databases like MySQL, MsSQL, Oracle, etc. It visually displays the current connection status and connection information of data sources and supports quick searches.

2) Data Table Management. The core function of the data reporting module, it enables users to import Excel spreadsheet data according to unified data specifications and standards. It also supports online data entry and editing (with real-time saving to the database, eliminating data loss concerns), as well as downloading data for local editing before uploading for unified storage management.

3) User Management. Primarily manages user permissions, allowing for the assignment of different table operation permissions (including data viewing, editing, and deletion) to different users. It supports multi-user collaboration by assigning different editing permissions (including CRUD operations) for different fields within the same table.

(3) Data Warehouse Management Module

This module comprises three sub-modules: data standard management, data model management, and data quality management.

1) Data Standard Management Module. Standardizes the structure, format, specifications, accuracy, and consistency of audit data. It includes functions such as standard search, code set management, and coding rule management. Standard search quickly matches relevant standard details within audit data standard information items by inputting keywords. Code set management unifies the management of various code sets required for audit data, which can be referenced by public attributes and data mart lists. Coding rule management manages automatically generated coding rules based on set rules, which can be bound to public attributes of audit data to detect data in entity tables and ensure compliance with audit data coding standards.

2) Data Model Management Module. The core objective of the integrated audit data management system is to bridge data silos and build a unified audit data warehouse that aligns with the standards of the education department by centralizing data from various systems with inconsistent data models. This module provides visual modeling capabilities for data models, adopting a combined top-down and bottom-up approach to model design, assisting in the design of standardized audit data models. It includes functions such as data model management, data model design, and entity table management. Data model management supports historical version management, enabling the process and transparency of the entire data model construction process. It facilitates the comparison of different versions, viewing change details, assisting auditors in managing different versions of models, and enabling the tracing of model design states at any node, thereby strengthening model control capabilities. Data model design ensures data consistency from the source by designing standardized data models based on audit data standards. Entity table management manages the entity tables generated by data models, displaying the current list of entity tables generated through data models.

3) Data Quality Management Module. To achieve comprehensive management and intelligent monitoring of audit data quality, this module allows for the setting of data quality inspection rules through "rule settings." These rules can then be applied and bound to data objects requiring quality checks for quality detection. It includes functions such as data quality rules and report management. Data quality rule management configures quality inspection rules for audit data, supporting the creation of various structural data quality inspection rules, including value range rules, enumeration rules, regular expression rules, and uniqueness rules. Data quality report management generates periodic data quality reports after quality inspections, presenting the data in vivid and multi-dimensional forms through diverse charts, helping departments gain a unified perspective on the quality of their respective data, identify optimization opportunities, and provide a basis and guidance for formulating quality improvement plans.

(4) Data Mart Module

The audit standard data warehouse serves as the core for centralized storage and integration of data, facilitating data fusion and sharing. This avoids redundant storage and processing of data across systems. Through the data mart module, audit-specific data marts can be created, primarily storing raw and derived data sourced from the standard data warehouse. These marts perform necessary processing according to the business requirements for audit data submission, providing the required result data and completing data submission. It encompasses functionalities such as data mart inventory, departmental data directory, and data openness management.

(5) Data Reporting Module

To ensure the accuracy, reliability of submitted data, and to cater to the statistical analysis needs of audit data, the data reporting module supports the visualization of reported data from the audit data mart through reports. It enables validation of related fields across reports, with mechanisms for alerting and flagging abnormalities. It also supports browsing, reviewing, and confirming report data. Furthermore, through reporting tools, multi-dimensional statistical analysis can be performed on audit data, making the data more dynamic and actionable for guiding relevant school decisions.

IV. HIGHLIGHTS AND CHALLENGES

At the end of 2020, the integrated audit data management system, while ensuring both security and accuracy, successfully supported 14 functional departments of Zhejiang University of Finance and Economics in completing the regular submission of approximately 12 million pieces of audit data information across 27 business systems, 59 tables, and nearly 900 items. As one of the first innovative pilot projects for digital reform in the education sector of Zhejiang Province, the integration of the integrated audit data submission system with the school's "New Generation of Comprehensive Data Center" has enriched the school's digital cross-scenario applications, further advancing its digital reform. This also provides reference ideas for the digital transformation of audit data management services in universities across Zhejiang Province.

1. System Highlights

1.1 Leveraging the "Data Hub" as the System Backbone

Essentially, the integrated audit data management system constructed by the university is not an independent, newly built system. Rather, to avoid the drawbacks of siloed management in a single system, the system's top-level planning is centered on leveraging the capabilities of the university's "data hub" – the comprehensive data center. By horizontally expanding the functionalities of the data hub platform, the system incorporates data entry and reporting modules. Building upon the university's prior achievements in comprehensive data governance, it realizes unified collection, standardized storage, secure retention, audit verification, centralized submission, and statistical analysis of audit data.

1.2 "Usage" Driving "Management"

In constructing the university's data infrastructure, alongside the foundation and management system, upper-level applications are also essential as user-facing entry points to enhance users' perception of data construction outcomes. Thus, the reporting of audit data to the Provincial Education Department, in a sense, provides a highly typical data application scenario, achieving a full-chain closed-loop management of audit data from "collection –

verification – standardization – storage – submission." This exemplary data application can, in turn, facilitate the improvement of the university's data management capabilities.

Simultaneously, this construction model can be quickly replicated and applied to other scenarios, with the university currently covering areas such as ideology and politics, party building, and student affairs.

1.3 Integration of Management and Service, Empowered by Digitalization

Through the construction of this system, the integration of management and service is achieved. Beyond data management tasks, it further enhances service capabilities. Utilizing a unified visualization platform, the system conducts multi-dimensional, in-depth, and visual analyses of audit data. By analyzing data quality and performing cross-departmental data correlation analyses, it "lets data speak the truth," enabling the university to promptly grasp the overall status of its operational processes, identify weaknesses in its business operations for improvement, and empower the digital reform of audit services with digitalization. This enhances the precise management and service capabilities, scientific decision-making level of the university's audit department, and modernizes the university's governance level and capabilities.

2. Difficulties and challenges encountered in the progress

During the overall advancement of the system construction, significant progress and breakthroughs have been achieved, yet concurrently, several difficulties and challenges have arisen, primarily in the following aspects:

2.1 Confirmation of audit data ownership departments

The audit data submission by the Zhejiang Provincial Department of Education encompasses data from 20 core business domains of the universities, including personnel, academic affairs, assets, student affairs, campus card, procurement, scientific research, finance, admissions, employment, continuing education, overseas management, campus traffic management, dormitory management systems, etc. The confirmation of ownership departments and systems for these data, encompassing which data originate from which department or system (or whether there is yet a management system for online data maintenance), offline data import and submission, and the designation of data rights and responsibilities, constitutes a highly complex task. It necessitates the establishment of a dedicated working group at the organizational level to coordinate departments across the university, thereby effectively advancing this endeavor.

2.2 Establishment of audit data standard model

The creation of an audit data standard model, under the university's existing "Data Hub" system and based on current school-level data standards, requires standardized field conversion and mapping according to the audit data standards prescribed by the Provincial Department of Education. This undertaking is immensely labor-intensive and necessitates the effective support of technical personnel and platform tools.

2.3 Horizontal expansion of "data hub" capabilities

The integrated audit data management system relies on the university's "Data Hub" for incremental development and iterative upgrades. The newly developed data submission and reporting modules must expand horizontally within the original "Data Hub" capability framework, not only fulfilling the requirements of audit data management but also maintaining consistency with the existing data system. This poses significant challenges to system development and construction, encompassing compatibility in product function design, user permission systems, data security design, and various other aspects.

V. CONCLUSION

In summary, this paper revolves around the construction of an integrated audit data management system, delving into the notable progress and groundbreaking achievements made during the system's advancement. Simultaneously, it provides an insightful analysis of the difficulties and challenges encountered. Through a series of measures, including confirming the ownership departments of audit data, establishing an audit data standard model, and horizontally expanding the capabilities of the "Data Hub," we have gradually established an efficient and

standardized audit data management system. Despite encountering numerous complex issues along the way, we have successfully overcome these obstacles through organizational structure optimization, technological investment, and multi-party coordination and cooperation, laying a solid foundation for the digital transformation of audit work. Looking ahead, with continuous technological advancements and optimized management, the integrated audit data management system will play an increasingly important role in enhancing audit efficiency, ensuring data quality, and promoting the modernization of educational governance.

REFERENCES

- [1] Yuan, J. J. (2021). Comprehensively promoting digital reform and striving to create significant landmark achievements for the "major window". *Policy Outlook*, (03).
- [2] Chou, L. Y., Du, T., & Lai, V. S. (2007). Continuous auditing with a multi-agent system. *Decision Support Systems*, 42(4), 2274-2292.
- [3] Dopuch, N., Gupta, M., Dan, A. S., & others. (2003). Production Efficiency and the Pricing of Audit Services. *Contemporary Accounting Research*, 20(1), 47-77.
- [4] Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, 58(5), 493-500.
- [5] Lan, G. S., Zhang, Y., Guo, Q., Kong, X. K., & Wei, J. C. (2020). Promoting digital transformation in higher education: Optimization, sustainability, and innovation—An interpretation and enlightenment of the "Top 10 IT Issues for 2020" report. *Open Education Research*, 26(5), 14-27.
- [6] Yang, H. Y. (2021). The development of auditing in developed countries and its enlightenment to the modernization of audit governance in China. *Northern Forum*, (02), 39-49.
- [7] Yu, P., & Li, Y. (2018). Research on data governance schemes in universities from the perspective of big data. *Modern Educational Technology*, 28(6), 60-66.
- [8] Zhao, D. X. (2019). Innovative development strategies for information management of college education in the era of big data. *Curriculum Education Research*, (10), 37+39.
- [9] Zhao, Y. X. (2009). Research on the whole process internal audit mode of construction projects in colleges and universities. *Science and Technology Management Research*, (04), 111-112.

About The Author: Linjun Kong , Director of the Digitalization Office at Zhejiang University of Finance and Economics, is a Senior Experimentalist with research interests in educational informatization, computer system software, data analysis, and network security.

Funding Project: Zhejiang philosophy and social sciences planning project "Research on the innovation and practice of university data governance models in the context of digital reform"(ID: 22NDJC108YB); Ministry of education's industry-university collaboration collaborative education project "research and application of university professional digital evaluation based on data middle platform" (ID: 231105921060349)



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details