Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research on Library Data Governance for Data Factorization

Yan Jiang*

Library of Zhongyuan University of Technology, Zhengzhou 450007, China

**Author to whom correspondence should be addressed.*

Abstract: Data factors are becoming the core driving force in the intelligent transformation of libraries. Based on a systematic review of the progress in data governance practices in libraries both domestically and internationally, this study delves into the mechanism by which data governance promotes data factorization and proposes implementation paths for data governance oriented toward data factorization. The aim is to facilitate the intelligent transformation and high-quality development of libraries.

Keywords: Data factorization; Libraries; Data governance; Mechanism of action; Practical paths

Online publication: December 16, 2025

1. Introduction

In 2022, the General Office of the CPC Central Committee and the General Office of the State Council issued the "Opinions on Promoting the Implementation of the National Cultural Digitization Strategy," proposing to "coordinate the construction of a national smart library system and establish a cultural digitization governance system." Following this, policy documents such as the "Opinions on Establishing a More Complete System and Mechanism for Market-Oriented Allocation of Factors" and the "Three-Year Action Plan for 'Data Factors ×' (2024–2026)" were released, initially forming a policy framework for data factors. Driven by the Digital China strategy and policies on data factorization, data has become a fundamental strategic resource for the country, reshaping the development models and value creation methods across various industries.

As an important distribution center for documentary information resources, libraries are witnessing data factors becoming the core driving force in their intelligent transformation. However, libraries generally face issues such as "data silos" and "information islands," resulting in low utilization rates of data resources (especially in market-oriented applications) and increased data security risks. In 2025, the National Development and Reform Commission and other departments issued the "Implementation Plan for Improving the Security Governance of Data Circulation to Better Promote the Market-Oriented and Value-Oriented Utilization of Data Factors,"

explicitly proposing to "establish a multi-party collaborative security governance system for data circulation by the end of 2027."

Data governance is being integrated into various fields of social governance from a new perspective to address data management challenges [1]. Library data governance is not only an inevitable choice for adapting to the digital transformation of education but also facilitates the effective development and innovative transformation of data value, thereby promoting the intelligent transformation and high-quality development of libraries.

2. Practical foundations for library data governance

Currently, libraries both domestically and internationally have embarked on a series of explorations and practices in data governance. Examples include the construction of data management and analysis platforms and smart library projects by institutions such as the National Library of China and Shanghai Library, as well as data open sharing initiatives by institutions like Harvard University Library and the British Library. These efforts drive the implementation of data governance.

2.1. Research on data governance practices in domestic libraries

In recent years, libraries in China have actively explored data governance in areas such as platform construction, standardization, and service innovation, forming distinctive practical models. The first phase of Shandong University Library's "Data Asset Platform" has been successfully applied in fields such as data governance, dimensional modeling, and data visualization, laying the foundation for data warehouse construction. The second phase will focus on integrating "Smart Library + AI Workshop + Data Large Model," leveraging large model technology to enhance cost-effectiveness in business management and services. Lanzhou University of Technology Library, in collaboration with Chongqing VIP Smart Map, has standardized the governance of approximately 549 million pieces of literature metadata. It has also harvested, transformed, cleaned, and integrated data on library space and seat utilization, establishing a user center to achieve refined management of different user groups. The National Library of China has developed unified data processing standards through ETL (Extract, Transform, Load) processes in its data management and analysis platform construction. Nanjing Library's big data platform has provided specialized data services for over 50 research projects and researchers, demonstrating its data-driven research support capabilities.

The data governance of domestic libraries exhibits notable characteristics: a shift from decentralized construction to platform-based and intensive development; an expansion from basic data cleansing to intelligent applications; and a transformation from single-institution governance to regional collaborative governance. However, challenges such as inconsistent data standards, inadequate privacy protection mechanisms, and insufficient in-depth data value mining persist.

2.2. Research on foreign library data governance practices

Foreign libraries have an earlier start in data governance, establishing relatively comprehensive institutional frameworks and technical systems, particularly accumulating rich experience in research data management, institutional repository construction, and privacy protection. According to a survey, 48 out of the top 50 university libraries in the United States offer open scientific data sharing services. Cornell University Library's "Data Repository" features tools capable of generating high-quality metadata in multiple formats, supporting research

data collection and sharing. The Research Councils UK (RCUK) has issued the "RCUK Common Principles on Data Policy," explicitly stipulating data management and usage policies and encouraging researchers to promptly make their research data publicly available. The Engineering and Physical Sciences Research Council mandates research institutions to provide at least a decade of data management services throughout the entire data lifecycle. Privacy protection has also formed a relatively complete policy framework and implementation mechanism. The American Library Association and the National Information Standards Organization have formulated relevant privacy management norms.

Libraries such as the University of London Library and the University of Oxford Library in the UK offer multiple data storage services tailored to different types of data and implement differentiated access controls based on data sensitivity. The construction of institutional repositories has recently shown a trend from traditional storage to intelligent curation. The "Data Repository" project at Cornell University Library, based on institutional repositories, aims to build a data curation platform and provide comprehensive services. The National Library of Greece, in collaboration with the Greek Open Knowledge Organization, has developed an entity management prototype to describe Greek-related entities and store them as knowledge graphs, providing a new solution for library data interoperability.

The data governance of libraries abroad exhibits the following notable characteristics:

(1) Institutions prioritize policy formulation, guiding data governance practices through clear policies and standards;

(2) A service-oriented approach is adopted, closely integrating data governance with research support and reader services;

(3) Technology-driven strategies are actively employed, leveraging emerging technologies such as linked data and knowledge graphs to enhance data value.

These experiences offer significant insights for data governance in Chinese libraries.

3. Mechanism of data governance

As a pivotal means in the process of data as a factor of production, data governance promotes the release of data value and the enhancement of library efficiency through systematic management mechanisms and technical operations. Together, they form an organic operational system.

3.1. Data rights confirmation and authorization mechanism

This mechanism serves as the legal foundation and institutional prerequisite for data as a factor of production. It is essential to establish hierarchical and categorized rules for rights management to lay the groundwork for compliant data circulation. The application of the "separation of three rights" is as follows: the right to hold data resources clarifies the library's legitimate position in holding original data such as bibliographic data and user behavior data; the right to process and use data grants libraries the authority to cleanse, annotate, and analyze original data; and the right to operate data products enables libraries to deploy processed data products for service applications or market transactions.

The data authorization mechanism acts as a bridge connecting data holders and users. For instance, digitized ancient book resources, special collection data, and academic resource metadata held by libraries, after undergoing desensitization and productization processing, can be authorized for use by research institutions and cultural enterprises through compliant channels. Specific measures to be noted during the process of data rights

confirmation and authorization include: establishing clauses for preserving public interests in data, requiring data users to fulfill certain public obligations; creating a mechanism for sharing data benefits, allocating a portion of data proceeds to public-interest data services; and introducing a data open licensing system to promote the lawful, orderly, and standardized processing and utilization of data.

3.2. Technology empowerment mechanism

This mechanism serves as the means and driving force for the realization of data as a factor of production in libraries. It is primarily manifested in several aspects:

(1) Data integration and processing technologies. These technologies standardize and normalize multi-source heterogeneous data. For instance, data middleware platforms, data lakes, and data warehouses provide efficient and secure environments for data integration and storage management. Based on data middleware platforms, a smart library data resource platform can be established, utilizing technologies such as distributed storage, metadata standards, and blockchain for evidence preservation. This platform collects data information on all business operations, management, and service objects within the library, harvesting, transforming, cleaning, and integrating the data to break down information silos across all business operations;

(2) Data analysis and mining technologies, such as machine learning and natural language processing. In data analysis, machine learning is employed to mine user behavior patterns and predict resource demands. With the aid of intelligent recommendation algorithms, personalized knowledge services are provided. This data-driven intelligent service model represents the future direction of library governance;

(3) Data visualization and interaction technologies. These technologies leverage RFID identification, artificial intelligence, and other technologies for interaction, using digital modeling and digital twins to achieve virtual displays. They intuitively present data analysis results through charts, dashboards, etc., supporting data-driven decision-making.

3.3. Security governance and risk control mechanism

Data security and privacy protection are prerequisites and baseline requirements for the realization of data as a factor of production in libraries. A security supervision system is a crucial measure for preventing data risks in libraries. Libraries should establish a comprehensive compliance rule system covering the entire lifecycle of data, from collection, storage, processing, and transmission to use and destruction, clearly defining the responsible entities and operational norms for each stage. Particularly for high-risk processes such as cross-border data flows and third-party data sharing, libraries need to develop specialized compliance guidelines and implement focused supervision.

Additionally, a preventive compliance system helps libraries proactively prevent data risks, significantly reducing compliance costs and legal risks associated with library data governance. Moreover, the rapid development of core technologies such as privacy computing, blockchain, and low-code development has made it possible to achieve “usable but invisible” and securely controlled data circulation and utilization. Trusted data space technology provides a solution for sharing sensitive data in libraries.

4. Implementation pathways

Against the backdrop of data as a key production factor, library data governance represents a systemic transformation from strategic planning to technological implementation, necessitating synchronized efforts across

multiple dimensions. The focus of action should extend beyond internal management to external services, with greater emphasis on data openness, shareability, and value enhancement.

4.1. Developing top-level design and strategic planning

Top-level design serves as the prerequisite and guarantee for the successful implementation of data governance. Specifically, it includes the following aspects:

(1) Formulating a data strategy as data governance should be integrated into the organizational strategic plan, aligning with the overall strategy of the library;

(2) Establishing modern governance concepts, clarifying the objectives and scope of governance, and grasping the direction and priorities of governance. The concept of openness and sharing forms the ideological foundation for the modernization of library data governance. Library leaders should actively cultivate an organizational culture of openness and sharing, exploring issues related to data openness, sharing, and value realization in new contexts such as data circulation and data transactions. This includes transitions from research-oriented open data platforms to open data platforms serving industry, academia, research, and application [2]. The scope of library data governance should gradually extend from data acquisition and sharing to data reuse and value addition;

(3) Establishing a Data Governance Management Committee, comprising decision-making, management, and execution levels from top to bottom;

(4) Formulating policies and regulations, establishing policies, standards, and guidelines for data management to regulate data governance activities;

(5) Designing governance processes, planning specific processes and methods for data governance to ensure orderly conduct of governance activities.

It is important to manage the relationships between different governance processes, between different stakeholders, and between data privacy and data openness and sharing, continuously adjusting and improving them through governance practice.

4.2. Building and improving the governance system

The construction of the governance system encompasses a multi-modal, end-to-end framework spanning from technology to application, strategy to management, and resources to services. Specifically, it can be divided into five dimensions (**Table 1**):

Table 1. Library data governance system

Layer	Core components	Implementation focus	Relevant standards & frameworks
Foundation layer	Data resources, technical infrastructure, human resources	Data resource inventory, trusted data space construction, talent acquisition & development	ISO/IEC 38505-1, GB/T 34960.5
Pillar layer	Data standards, data quality	Metadata standards, data quality assessment	ISO 8000, GB/T 36073-2018
Safeguard layer	Organizational structure, policies & regulations, governance processes	Governance committee establishment, management system development, process design	COBIT, ITIL, GB/T 34960.1
Core layer	Data collection, storage, processing, analytics & applications	Data lake construction, ETL processes, analytical model development	ISO/IEC 20547, GB/T 38667
Protection layer	Data security, privacy protection, compliance management	Access control, encryption & masking, compliance auditing	

4.2.1. Foundation layer

This layer encompasses fundamental elements such as data sources, technological infrastructure, and human resources. Governance of data sources begins with a comprehensive inventory and cataloging process, clarifying which data is available, who is utilizing it, and how it is being used. Data infrastructure represents a new type of infrastructure that integrates hardware, software, and standard specifications. It spans five domains: access, storage, computation, management, and data enablement. Its construction includes technical facilities such as trusted data spaces and data arenas. The development of human resources should revolve around a trinity framework of "data literacy + job role restructuring + collaborative mechanisms" to meet emerging service demands.

4.2.2. Pillar layer

This layer includes components such as data standards, data quality, and data security. The library data standards system should encompass metadata standards, master data and reference data standards, data exchange and interface standards, and data modeling standards. Data quality management involves establishing a data supervision mechanism, forming a data quality management system, monitoring and disclosing data quality issues, and providing detailed query capabilities for issues and suggestions for quality improvement. This, in turn, promotes data transparency and facilitates data sharing and collaboration among libraries. Data security encompasses processes such as hierarchical and categorized data management, role- or attribute-based access control, encryption and de-identification, security auditing, and monitoring.

4.2.3. Support layer

This layer consists of supporting elements such as organizational structure, policies and regulations, and governance processes. To establish a collaborative governance mechanism, a typical three-tier organizational structure should be constructed within the collaborative governance framework of "market + government + public libraries" [3]. Policies and regulations include data management policies, data quality standards, data security norms, and data access permission control rules. This involves the formulation and refinement of standards such as data preservation and conversion norms, data ethics guidelines, technical operation standards, and data security protection rules [4]. The governance process standardizes governance activities, with key processes including data change management, data issue resolution, data request and authorization, and compliance auditing.

4.2.4. Core layer

This layer encompasses core activities such as data collection, storage, processing, analysis, and application. It describes the entire process of data flow and value appreciation within the library. The data middleware, serving as a centralized data management platform, supports the construction of various productivity tools from data collection to processing and service output. These tools include big data development suites, data quality management, data models, data API platforms, tag factories, data lineage, reporting engines, and visual dashboard engines, among others [5].

4.2.5. Protective layer

This layer includes protective measures such as data security, privacy protection, and compliance management. It is essential to provide various data security strategies, including data encryption and desensitization, obfuscation, and account monitoring. Privacy protection and compliance management must strictly adhere to relevant laws, regulations, industry standards, and policy requirements, ensuring that all operations are conducted within the legal

framework through supervision.

4.3. Ecosystem construction and service model innovation

The ultimate goal of library data governance is to create a self-sensing, self-deciding, and self-optimizing service ecosystem. In terms of vertical development, libraries integrate deeply with publishers, resource providers, and others to establish new platforms, enabling business alignment and the co-construction and sharing of data. Horizontally, libraries will integrate data resources and services from various channels to create new composite information data services. For instance, library collections can be transformed into a knowledge repository through matrix construction using vector databases, graph databases, and structured databases, providing knowledge services to users through services such as retrieval-augmented generation and intelligent agents [6]. Additionally, leveraging metaverse technology, libraries can offer users scenario-based virtual tours, deep interactive experiences, and immersive reading [7].

5. Conclusion

As data becomes a key production factor in the digital era, libraries face both unprecedented opportunities and significant governance challenges. This study systematically reviewed domestic and international practices, clarified the mechanisms through which data governance drives data factorization, and proposed implementation pathways spanning top-level design, system construction, and service innovation. The findings indicate that effective data governance not only resolves issues such as data silos, inconsistent standards, and security risks but also enables libraries to unlock the value embedded in their data resources, enhance intelligent services, and support high-quality development. Looking forward, libraries should continue to strengthen policy coordination, accelerate technical innovation, and build open, collaborative data ecosystems. By advancing toward refined governance and value-driven utilization, libraries can more fully realize the transformative potential of data factorization and better support the intelligent development of cultural and knowledge services.

Disclosure statement

The author declares no conflict of interest.

References

[1] Hu H, Lou C, 2019, Government Data Governance in the Context of Data Openness: Theoretical Logic and Practical Directions. Information Studies: Theory & Application, 42(7): 41–47.

[2] Sun J, Li Y, 2024, Research on Data Elements from the Perspective of Information Resource Management Discipline. Documentation, Information & Knowledge, 41(2): 6–12.

[3] Hu T, 2024, Research on the Mechanism and Path of Data Value Realization in the Digital and Intelligent Service Transformation of Public Libraries. Journal of the Sichuan Library, 262(6): 54–61.

[4] Sun H, 2022, Main Issues, Conceptual Framework, and Practical Paths of Public Data Governance in Libraries. Research on Library Science, 2022(3): 11–17.

[5] Pan Y, Zheng F, 2020, Research on the Risk Assessment Mechanism of Data Governance in China's Public Libraries from the Perspective of Administrative Law. Research on Library Science, 2020(22): 20–26.

[6] Jin J, 2025, AI-Driven Paradigm Reconstruction and Innovation Paths for Future Libraries. Journal of Library Science in China, 2025: 1–15.

[7] Yan E, 2025, Mutual Empowerment Between New Quality Productivity and the Metaverse Library: Realistic Dilemmas and Development Paths. Library and Information Service, 69(17): 16–26.