
AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202506.00006

Data Empowerment and Smart Reconstruction: Applications and Future Prospects of Big Data Technology in Public Libraries

Authors: Wu Wenyong

Date: 2025-06-01T18:04:22+00:00

Abstract

At the current juncture where the digital economy development wave intersects with the implementation of the cultural powerhouse strategy, the public library sector is undergoing a paradigmatic transformation from traditional “document resource centers” to modern “knowledge service hubs.” Big data technology has rightfully emerged as the core engine driving this shift. This study synthesizes case analysis and normative research methodologies to systematically examine the practical applications of public libraries in optimizing management decision-making, innovating resource integration, and reconstructing service models; across three dimensions—management efficacy enhancement, service quality upgrading, and resource construction transformation—its value empowerment mechanisms are profoundly elucidated.

Full Text

Data Empowerment and Intelligent Reconstruction: The Application of Big Data Technology in Public Libraries

In the current era where the digital economy wave intersects with the national strategy of building a strong cultural nation, the field of public libraries is undergoing a paradigmatic transformation from traditional “document resource centers” to modern “knowledge service hubs.” Big data technology stands as the core engine driving this shift. This paper synthesizes case study and normative research methods to systematically analyze the practical applications of public libraries in optimizing management decisions, innovating resource integration, and reconstructing service models, while deeply revealing the value-enabling mechanisms across three dimensions: management efficiency improvement, service quality upgrading, and resource construction transformation.

To address real-world challenges such as insufficient data governance systems and bottlenecks in technological application capabilities, we propose solutions including the construction of standardized data frameworks, the application of privacy computing technologies, and the urgent strengthening of composite talent cultivation. Looking ahead, intelligent service scenario construction holds broad prospects, precise decision support deepening contains great potential, integrated ecological co-construction shows multiple possibilities, and inclusive service upgrading is full of development opportunities—these represent future trends under the dimension of technological integration.

Our research concludes that big data technology functions not merely as an efficiency tool but as a strategic element that significantly reshapes the value positioning of libraries. A new public cultural service ecosystem featuring data intelligence as its core characteristic must be realized through systematic transformations including improvements to data governance systems, optimization of organizational capability structures, and innovation in service models.

Keywords: Big data technology; Smart library; Data governance mechanism; Service innovation path; Public library transformation

1. Introduction: The Transformation Logic of Public Libraries in the Big Data Era

As the digital economy now accounts for over 41.5% of GDP (according to 2024 data from the China Academy of Information and Communications Technology), public libraries, as social knowledge infrastructure, face development challenges including lagging resource digital transformation, inefficient matching of service supply and demand, and insufficient decision support capabilities. The technical empowerment system built upon big data's 5V characteristics—Volume, Velocity, Variety, Value, and Veracity—provides a new path to resolve these difficulties. Through deep mining of user behavior data, resource utilization data, and equipment operation data, libraries can achieve management transformation from experience-driven to data-intelligence-driven models. By leveraging cross-system data fusion and knowledge graph construction, they can upgrade resource services from discrete supply to networked associative supply. Relying on intelligent analysis of real-time data streams, they can promote service model transformation from standardized services to scenario-based precise services. This paper systematically analyzes the internal mechanisms through which big data technology reshapes the public library ecosystem, based on deep coupling between technological applications and business scenarios, aiming to provide theoretical reference and practical guidance for smart library construction.

2. Application Status and Typical Practices of Big Data Technology in Public Libraries

2.1 Data-Driven Management Decision Optimization

2.1.1 Innovative Practice of Intelligent Library Management Platforms A provincial library, in collaboration with a technical team, developed a “Smart Library” management platform that constructs a data middle-platform system based on microservices architecture, achieving organic integration of multi-source heterogeneous data including business management systems, IoT sensor data, and user behavior logs. By deploying over 200 intelligent perception terminals, the system collects 128-dimensional data in real time, covering building energy consumption, equipment operation status, and reader movement trajectories, and uses digital twin technology to construct a visual model of library operations. The management cockpit system has achieved remarkable results, increasing equipment maintenance efficiency by 32% and optimizing space resource utilization by 28%, forming a closed-loop management mechanism of “data collection—intelligent analysis—decision execution—effect feedback.” This project has been recognized by the Ministry of Culture and Tourism as a “Demonstration Case of Public Culture Digital Innovation.”

2.1.2 Dynamic Reader Profiling and Precision Service System Construction Based on K-means clustering algorithms and natural language processing technology, a dynamic reader profiling system has been constructed that integrates borrowing records (40%), digital resource usage logs (30%), offline activity participation data (20%), and social platform interaction information (10%), encompassing 36 tags including basic attributes, reading preferences, and knowledge needs. A pilot library identified seven core user groups through hierarchical clustering. For “STEM education demand” parent groups, it pushed combined services of “science picture books + maker activities,” increasing related resource utilization by 65%. For the “elderly digital divide” group, it customized large-font interfaces and voice interaction functions, increasing digital resource visits from this group by 40%. The service model has shifted from “experience-based universal supply” to “data-driven precise adaptation.”

2.2 New Explorations in Resource Coordination and Service Forms

2.2.1 Cross-Domain Information Fusion and Knowledge Discovery Support Through comprehensive ETL (Extract, Transform, Load) processes, multi-source information including library OPAC catalogs, CNKI academic databases, JSTOR foreign language materials, and MOOC course sections has been reorganized to construct a cross-disciplinary concept network with over 100,000 entity nodes and more than 500,000 relational links. Driven by entity recognition and link mining, the system reveals latent knowledge about author collaboration networks, research topic evolution paths, and technological innovation clues. For instance, when searching with keywords like “artificial intelligence,” the system simultaneously outputs core collection books (32

volumes), approximately 200 recent advanced papers, and 15 organized domain lecture videos with specific location navigation in the library. This creates a multi-dimensional access channel of “knowledge point—link—extended connection structure,” improving retrieval efficiency by over 40% in practice.

2.2.2 Data-Driven Innovation in Reading Promotion Scenarios A refined recommendation model based on the “3A5-step” framework has been proposed, referencing a high-correlation data graph composed of user interaction logs (Activity), social circle mapping (Association), and real-world context (Ambience), implemented across five stages: “demand perception—targeted recommendation—multi-channel connection—effect tracking feedback—next-round strategy fine-tuning.” For Z-generation reader groups, the system combines internet check-in frequency (high activity rate between 7-10 PM) with real-time space heat analysis (creative zones account for 35%), launching innovative reading activities such as “night theater-style shared reading” and “AR space treasure hunting.” Participation numbers increased by 3.2 times compared to traditional promotional posters. Each activity’s effectiveness feeds back into the overall demand perception architecture, solidifying a virtuous cycle of “data-based—experience optimization—continuous improvement.”

3. Core Value of Big Data Technology Empowering Public Libraries

3.1 Management Efficiency Improvement: From Experience-Based to Data-Intelligent Decision Making

A library operations indicator system has been constructed encompassing 22 core indicators across three dimensions: resource management, service efficiency, and user experience, enabling “digital twin” management decisions through real-time data dashboards. Equipment IoT systems based on LSTM neural networks predict failure probabilities, transforming traditional post-failure maintenance into preventive maintenance, reducing maintenance costs by 38%. Reader flow prediction models combine external variables such as weather data and school holiday schedules to dynamically adjust regional opening hours and staffing configurations, increasing service response speed during peak hours by 25%. The data middle-platform breaks down departmental data barriers, achieving digital integration of procurement, circulation, and evaluation business processes, and promoting flat, agile organizational management.

3.2 Service Quality Upgrading: From Scale Supply to Value Co-Creation

Big data technology reconstructs service interaction paradigms. At the demand insight level, text mining technology analyzes 100,000 reader consultation records, identifying high-frequency needs such as “thesis writing guidance” and “patent retrieval,” and developing specialized resource packages that in-

crease related service utilization by 55%. At the service execution level, real-time data flow dynamically adjusts service configurations—for example, automatically opening backup reading areas when seat usage data shows vacancy rates below 20%, and triggering intelligent push mechanisms for “e-book reading gift packages” during extreme weather. A provincial library’s user satisfaction survey shows that big data-driven services increased overall reader satisfaction from 72.3% to 88.6%, shifting the service value creation model from “collection-centered” to “reader service-centered.”

3.3 Resource Construction Transformation: From Extensive Procurement to Precise Allocation

A resource procurement decision model has been established incorporating 18 parameters including borrowing volume, reservation rate, user ratings, and academic influence, using collaborative filtering algorithms to predict resource needs in specific fields. Taking social science books as an example, the system automatically generates procurement priority lists by analyzing three-year borrowing data (15% annual growth), Douban book ratings (60% with scores \$ \$8.5), and author influence indices, increasing utilization of such books from 62% to 85%. In digital resource procurement, sentiment analysis technology evaluates user comments to optimize database procurement combinations, increasing the renewal rate of frequently used databases such as CNKI by 20% and improving the funding input-output ratio by 35%.

4. Challenges and Solutions in Big Data Technology Application

4.1 Data Governance Dilemma: Three Dimensions of Standards, Security, and Quality

4.1.1 Construction of Data Standardization and Sharing Mechanisms

To address interoperability challenges caused by heterogeneous business systems, we propose establishing a three-tier “national—provincial—municipal” data standard system: the foundation layer adopts DC metadata standards for unified resource description, the application layer formulates API interface specifications (such as RESTful standards) to ensure system interoperability, and the security layer establishes data classification and management methods (referencing ISO 27001 standards). Promote the construction of cross-regional library data sharing platforms, piloting the “Yangtze River Delta Public Library Data Exchange Center” to achieve heterogeneous data interoperability among 200+ libraries, with a target data sharing rate above 70%.

4.1.2 Technical Solutions for Privacy Protection and Compliant Utilization

Construct a full lifecycle security protection system covering “data collection—storage—use—destruction”: adopt federated learning technology at the data collection end to achieve privacy computing with “data stationary,

model mobile”; implement encryption (AES-256 algorithm) and access control (RBAC role-based permission management) during storage; apply differential privacy technology to add -differential noise ($=0.5$) during usage to ensure anonymization of user profile data. Establish a data compliance review committee to conduct regular assessments of GDPR and Personal Information Protection Law compliance, ensuring legal and compliant data utilization.

4.2 Technology Application Bottlenecks: Capacity Building and Sustainable Development

4.2.1 Construction of Composite Talent Cultivation System Implement a “dual-wheel drive” talent strategy: internally, through programs such as “Smart Library Workshops” and “Data Analyst Certification Training,” achieve basic data processing capabilities for 30% of librarians within three years; externally, co-build interdisciplinary master’s programs in “Library Science + Data Science” with universities such as Wuhan University and Nanjing University to cultivate composite talents with domain knowledge and technical skills. Establish a “Data Service Specialist” position sequence, constructing a career development path from “Document Service Position—Data Processing Position—Knowledge Service Position” to promote the transformation and upgrading of librarians’ capability structures.

4.2.2 Sustainable Models for Infrastructure Construction A sophisticated technological system of “hybrid cloud collaboration + edge intelligent computing” is currently employed to reduce overall operational expenses. Critical data such as reader service management and material procurement are deployed on local private cloud platforms, significantly enhancing information security. Meanwhile, historical borrowing records from ten or more years ago are migrated to public cloud architecture, reducing storage costs by an average of approximately 40%. Additionally, a “multi-channel investment model” integrating government funding with social capital is being explored and promoted. Consequently, a resilient and progressively evolving new technological ecosystem is gradually forming, establishing a solid foundation for continuous technological application expansion.

5. Future Development Trends Driven by Big Data Technology

5.1 Intelligence: Human-Machine Collaboration Reconstructing Service Scenarios

Future planning for public libraries focuses on creating sophisticated comprehensive information perception environments rather than single spatial structures. For instance, ultra-high frequency RFID-based collection inspection robotic arms operate at approximately 5,000 books per hour. When combined with visual image analysis algorithms, they can subtly complete operations such as

book repositioning and classification screening with accuracy maintained around 99.5%. Additionally, intelligent inquiry partners under multi-modal data integration have been developed, accessing modules such as automatic speech recognition (ASR), natural language understanding (NLU), and knowledge network inference technology, making $24\$ \times \7 uninterrupted Q&A nearly a reality with typical response times under ten seconds. For environmental regulation, pressure monitoring elements and emotion interpretation camera devices are simultaneously introduced to dynamically control readers' psychological states. Lighting area brightness adjustment is set to one-tenth increments, while ambient background noise is maintained at approximately 40-50 decibels, creating a personalized immersive experience atmosphere.

5.2 Precision: Full-Process Digital Twin Decision Support

A virtual twin system covering three major links—resource allocation, service supply, and effectiveness feedback—is gradually being constructed. This sophisticated system relies on real-time information mapping for multi-scenario推演. For example, in document procurement scenarios, the platform combines historical borrowing weight (approximately 60% of the total), disciplinary direction trends (quantitative adjustment, around 20%), and budget constraints (preset 20%), using a genetic algorithm-like strategy model to make intelligent recommendations for suitable procurement book combinations. Consequently, management decision-making speed has nearly doubled year-over-year. In activity planning, introducing discrete event dynamic simulation tools such as AnyLogic allows forward assessment of user participation, related material consumption, and reader feedback, facilitating real-time optimization of execution details. This establishes a new operational model of “physical library space—digital mirror entity—intelligent decision-making” continuous loop regulation.

5.3 Integration: Cross-Domain Ecosystem Co-Construction and Value Co-Creation

Industry boundaries are quietly broken, forming an extremely complex big cultural data growth circle. Through close collaboration with museums and science exhibition venues, a cultural big data alliance is jointly planned. In this process, over 3 million pieces of digital cultural relic information resources have been integrated, and more than 50,000 science popularization short videos have been synchronized. When involving cross-domain content, multi-dimensional entity network graph hierarchical relationships become more apparent, with correlation connections rising to approximately 40% of the original level. Meanwhile, diverse enterprises such as ByteDance and Tencent have participated in cooperation, launching a massive “Data Information Reading Cloud Service Platform” plan. Relying on large-scale cloud computing architecture, basic storage costs have been reduced by nearly 30%, while edge computing has accelerated user access experience speed by an average of 20%. Urban public culture connects with market mechanisms, significantly enhancing overall coordination capabil-

ties. Furthermore, deep involvement with the “Urban Intelligent Governance Core System” enables continuous sharing of social data such as population flow and education supply-demand. Consequently, new control support for cultural supply within urban scope has initially become apparent, with the public accessing local culture in a more flexible and free manner.

5.4 Inclusiveness: Data Equity and Barrier-Free Service Upgrades

To alleviate the digital divide, constructing an inclusive service system covering diverse objects becomes a comprehensive information platform design task. For visually impaired user groups, it is necessary to develop intelligent navigation assistance devices based on acoustic signal position detection mechanisms, combined with human voice dialogue interaction technology, to subtly achieve “one-click access to Braille resources plus audiobooks.” For elderly readers, interface experiences should be extremely simple and intuitive. Through trajectory and habit data, the total operation process is repeatedly adjusted and streamlined, reducing the number of function option entrances by nearly half. Additionally, for seniors who do not frequently use network devices, “elderly-friendly digital environment” modules such as telephone appointment and in-library agency services are launched. Meanwhile, emerging perceptual computing technologies can sensitively detect individuals in information anxiety states, automatically activating “one-on-one knowledge guidance” special assistants to ensure equal coverage of public resources and benefit all social strata. This setting reflects comprehensive care attributes, invisibly promoting the entire intelligent interaction ecosystem toward a more comprehensive and practical shared development direction.

6. Conclusion: Toward a New Library Ecosystem Driven by Data Intelligence

Currently, complex data intelligence technology is quietly permeating various public libraries, gradually expanding from functional utilization of single operational tools to entirely new changes in service methods and operational logic structures. Actual conditions show that not only has conventional management level been increasingly improved, but various new service forms have continuously emerged. More profound changes are reflected in the systematic adjustment of value generation chains, shifting from initial emphasis on literature resource collection to knowledge demand orientation, with institutional self-interest no longer occupying the primary position and community users becoming the core of overall design. In the long run, to continuously obtain development momentum, focusing solely on shallow breakthroughs in technological application links can hardly meet future expectations. Current academic and industry attention focuses on how to construct an overall pattern where a comprehensive data governance framework, underlying digital information platform foundation, and flexible multi-dimensional talent system complement each other.

In long-term discussions within this field, numerous measures have been summarized. Industry standard establishment, data-driven concept promotion, and high-quality human resource cultivation systems all represent critical steps that must be promoted. Through subtly strengthening these aspects and continuously advancing organizational mechanism optimization, a unique, transparent, and fair information security assurance environment can be formed. Ultimately, a new public reading space ecosystem is gradually established, characterized by cultural content sharing and dissemination relying on data intelligence capabilities.

This trend seems not only related to self-growth processes and social knowledge acquisition efficiency but also inseparable from the foundational guarantee of a mutually beneficial high-quality spiritual life for all. Consequently, under the guidance of macro goals such as learning society cultivation and mass cultural power consolidation, related efforts possess strong practical significance and strategic importance.

References

- [1] Chu Jingli, Zhang Dongrong. The path and strategy of smart library construction [J]. Journal of National Library of China, 2020, 29(01): 5-12.
- [2] Meng Xuemei, Wang Mangang. Research on the construction of user portraits in public libraries under the big data environment [J]. Library Forum, 2021, 41(03): 105-113.
- [3] Zhang Xiaolin. The development direction of libraries in the era of big data [J]. Journal of Chinese Libraries, 2019, 45(02): 4-12.
- [4] Chen Yaosheng, Li Jing. Research on library cross-system data fusion based on Knowledge Graph [J]. Intelligence Theory and Practice, 2022, 45(05): 118-123.
- [5] Huang Ruhua, Li Baiyang. Research on data privacy protection in library services [J]. Library and Information Work, 2021, 65(14): 5-13.
- [6] Ministry of Culture and Tourism. “14th Five-Year” Public Cultural Service System Construction Plan [Z]. 2021.
- [7] Gansu Provincial Library. White Paper on the Construction of Smart Gantu Comprehensive Management Platform [R]. 2023.
- [8] China Library Society. Report on the Development of Big Data Applications in Public Libraries (2022) [Z]. Beijing: National Library Press, 2022.
- [9] GB/T 36344-2018 Evaluation Indicators of Library Service Effectiveness [S]. Beijing: China Standard Press, 2018.
- [10] Personal Information Protection Law of the People’s Republic of China [Z]. 2021.

[11] World Library and Information Congress (IFLA). White Paper on Big Data and Library Service Innovation [EB/OL]. (2020-10-15).

[12] Shanghai Library. Innovation Practice Cases of Reading Promotion Based on Big Data [EB/OL]. (2023-08-20).

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.