Building High-Quality Expert Database Datasets: Guangdong's Practices in Leading AI Empowerment and the Development of a Unified National Market

Jiong Zheng[1], Haowei Qiu[2], Lihua Huang[1*]

[1] Guangdong Digital Government Affairs Association, Guangzhou, 510030, China

[2] Guangdong Provincial Government Service and Data Affairs Center, Guangzhou, 510030, China

* 410918458@qq.com

https://doi.org/10.70695/IAAI202504A14

Abstract

Against the backdrop of accelerating the development of a unified national market, standardization and digitization in the field of public resource transactions are critical steps. As a core element of bidding and tender activities, the data quality of the evaluation expert database directly impacts the efficiency and fairness of resource allocation. This paper takes the innovative practices of Guangdong Province's comprehensive evaluation expert database data specification as a case study, delving into how systematic data governance can be employed to construct high-quality expert datasets. These datasets support application scenarios such as AI-driven random selection and evaluation assistance, thereby providing a solid foundation for building a high-standard and high-efficiency unified national market.

Keywords Expert Database; Data Specification; High-Quality Dataset; Artificial Intelligence

1 Introduction

Currently, China is committed to dismantling various forms of market fragmentation, advancing the development of a unified national market, and fostering a new development paradigm. The public resource trading market is a vital component of the unified national market, and its healthy operation relies on fair, impartial, and efficient bidding mechanisms. Within this framework, evaluation experts play the crucial role of "referees", whose professional competence and impartiality directly impact the quality of transactions.

However, traditional management of evaluation expert databases commonly faces challenges such as inconsistent data standards, uneven information quality, and prevalent data silos. These issues not only hinder the cross-regional sharing of expert resources but also impede the effective application of advanced artificial intelligence technologies. At the core of AI applications lies "data fuel"—low-quality, non-standardized data cannot train reliable models and may even lead to negative outcomes characterized by "garbage in, garbage out". Therefore, constructing a high-quality, standardized dataset for evaluation expert databases is a prerequisite for enabling comprehensive sharing of expert resources, realizing AI-driven intelligent applications, and ultimately supporting the development of a unified national market.

Guangdong Province has taken the lead by introducing the nation's first provincial-level data specification standard for a comprehensive evaluation expert database. This initiative provides systematic, digitalized, and standardized support for the intelligent management of expert databases and the smart supervision of bidding activities. It offers a valuable practical example of how building high-standard expert database datasets can serve the broader goal of constructing a unified national market.

2 Background

Recently, the Recommendations of the Central Committee of the Communist Party of China for Formulating the 15th Five-Year Plan for Economic and Social Development, reviewed and adopted at the Fourth Plenary Session of the 20th Central Committee of the Communist Party of China, identifies "deepening the development of a unified national market and continuously leveraging the advantages of an ultra-large-scale market" as one of the primary objectives for economic and social development. It

explicitly calls for "resolutely dismantling obstacles and bottlenecks hindering the development of a unified national market", improving systems such as property rights protection, market access, information disclosure, social credit, mergers and acquisitions, and market exit, eliminating barriers in areas including factor acquisition, qualification recognition, bidding and tendering, and government procurement, and regulating the economic promotion activities of local governments [1].

According to Measures for the Administration Bid Evaluation Experts and Bid Evaluation Expert Databases (Order No. 26 of 2024 of the National Development and Reform Commission), an evaluation expert database refers to an electronic information system that stores evaluation expert information and provides functions such as selecting experts for evaluation, assisting the management of units responsible for establishing expert databases, and offering necessary services to evaluation experts. It plays a critical role in the development of a unified national market, particularly in safeguarding fairness and impartiality in areas such as factor acquisition and bidding.

For a long time, there has been a gap in the standardization of data for comprehensive evaluation expert databases both in Guangdong Province and nationwide. Irregularities in information management at the operational level of bidding activities have been prevalent, urgently necessitating solutions to current management challenges caused by incomplete and inaccurate expert data.

In accordance with the Opinions of the General Office of the State Council on Innovating and Improving Systems and Mechanisms to Promote Regulated and Healthy Development of Tendering and Bidding Market (State Council Document No. 21, 2024) and other related policy directives, efforts are being made to optimize the full-cycle management of evaluation experts, promote the integration of bidding activities with new technologies such as big data, cloud computing, artificial intelligence, and blockchain, and enhance the sharing of expert resources for remote and cross-regional evaluations.

Guiding the governance of existing expert data and the development of future systems through scientific and unified data specification standards for comprehensive evaluation expert databases represents a crucial step in implementing national directives and the requirements of provincial party committees and governments. It serves as a solid foundation for supporting the high-quality development of evaluation and bidding-related work across the province and stands as a core element in leading industry technological advancement through standards and better leveraging the safeguarding role of standards.

3 Building High-Quality Datasets Through Standardization

During the 16th thematic study session focusing on "strengthening the guiding and safeguarding role of standards, and promoting high-quality economic development through standard enhancement", the State Council emphasized that "standards serve as a crucial foundational system and play an important role in constructing a modern industrial system and building a unified national market" [2]. The expert database of Guangdong Province supports the full-cycle management of over 40,000 experts across the province, facilitates the evaluation of major public resource trading projects, and enables the sharing of expert resources for remote and cross-regional evaluations. At the same time, it must coordinate with data sharing for the provincial Development and Reform Commission's regulatory platform, as well as inspections and investigations by administrative supervisory authorities, discipline inspection and supervisory organs, judicial authorities, and audit departments. The quality of expert data is a core element underpinning high-quality evaluation processes, and its significance is self-evident.

Guangdong's pioneering development of a data specification for the comprehensive evaluation expert database is not merely a technical standardization effort but a systemic governance project encompassing the entire data lifecycle. By conducting assessments of data quality, mobilizing industry expertise to formulate data standards, and guiding data governance and system development, it comprehensively enhances the quality and service capacity of the expert database. This represents a significant practical exploration in building high-quality datasets through standardization to promote high-quality economic development.

3.1 Conducting Data Quality Assessment

During the initial phase of standard formulation, research and analysis were conducted on system operations, database architecture, and related data issues. Data queries and data analysis were performed on the expert database platform's tables across dimensions such as compliance, completeness, accuracy, consistency, and timeliness [3]. Data-related issues were further validated through historical business

transaction records and customer service hotline feedback, supplemented by data analysis based on business scenarios including expert onboarding, renewal applications, and routine assessments.

First, compliance refers to assessing whether data meets the requirements of data standards, data models, business rules, metadata, or authoritative reference data, as well as security specifications, including data permission management and data masking.

Second, completeness involves evaluating whether values are assigned according to data rules, mandatory fields are filled, conditional mandatory fields are completed as required, and whether data is missing for specific time periods.

Third, accuracy entails assessing whether data formats—including data types, value ranges, data lengths, and precision—meet expected requirements, as well as the repetition rate of specific fields or datasets and the occurrence rate of dirty data.

Fourth, consistency involves evaluating the uniformity of the same data stored across different locations, ensuring that when data changes, the same data stored in multiple locations is updated synchronously, and verifying the consistency of related data based on predefined constraint rules.

Fifth, timeliness refers to assessing the extent to which time-stamped record counts, frequency distributions, or delay times align with business requirements.

3.2 Developing Data Standards Scientifically

Based on the outcomes of the preliminary data governance assessment, a comprehensive and scalable governance model—ranging from data specifications to security management—has been established through organized policy research, systematic investigations, and standard development efforts within the industry.

Clearly Defined Data Quality Dimensions for Assurance

The core innovation of the specification lies in concretizing data quality requirements into five actionable and verifiable dimensions. Compliance is achieved by referencing 13 national and industry standards, which unify the definitions, formats, and value ranges of over 300 data fields, thereby providing a "common language" for data interoperability. Completeness requires clarifying the mandatory status of each data element to ensure comprehensive expert profile information, offering full-feature dimensions for AI models. Accuracy entails strictly specifying data types, lengths, precision, and uniqueness, eliminating ambiguous or erroneous information at the source and safeguarding the reliability of AI analysis and decision-making. Consistency involves establishing logical constraints and synchronization mechanisms between data to resolve contradictions between historical and current data caused by multiple system upgrades and iterations. Timeliness requires defining data update cycles and temporal relationships to ensure that AI systems utilize dynamic and up-to-date expert information. These five dimensions together form a rigorous data quality framework [4], laying a solid foundation for generating "high-quality fuel" applicable to AI training and applications.

Full Lifecycle Data Classification Management

The specification categorizes expert data into three major types, covering the entire process from "entry into the database-retention in the database-exit from the database". Basic Information includes expertise, professional achievements, and research outcomes, among other details [5]. It is maintained by experts themselves and serves as the foundation for AI-generated professional competency profiles. Management Information covers training and education, annual evaluations, and dispute handling, among other aspects [6]. It dynamically records an expert's development and compliance status, providing critical evidence for AI to assess expert reliability. Evaluation Participation Information consists of project history, evaluation performance, and assessment results, among other records [7]. It forms an expert's "digital performance file" and serves as the core data source for AI-enabled intelligent matching and risk warning. Such classified management transforms data from static archives into computable resources that dynamically reflect expert status, providing structured data support for AI application scenarios.

Multi-layered Data Security Safeguards

The specification establishes multiple safeguards to fortify the security of expert data. Legality and compliance are ensured by strictly adhering to regulations such as the Cybersecurity Law, guaranteeing lawful collection, storage, and use of data. Field-level encryption is implemented, with sensitive information encrypted using state cryptographic algorithms at the field level. Minimum permission control enforces the principle of least privilege, prohibiting any querying or modification unrelated to work requirements. Data retention requires electronic records to be permanently preserved, dynamically updated, and fully traceable throughout their lifecycle [8]. By implementing these multi-layered data

security measures, expert privacy is protected, data security is ensured, and a reliable foundation is provided for AI training.

3.3 Strengthening Standard Referencing and Promotion

The achievements of Guangdong Province in innovatively developing the data specification for the expert database are reflected not only in the formulation of the specification document itself but also, more crucially, in ensuring the practical effectiveness of the data specification through strengthened referencing and promotion of the standard. By reinforcing the referencing and promotion of the expert database data specification, Guangdong Province has not only addressed practical issues in the construction of the expert database but also laid a solid foundation for its long-term development. Standard referencing ensures the enforceability of the specification, while widespread promotion fosters understanding and adoption. The combination of the two makes the expert database data specification a truly effective tool for advancing the management level of expert databases and provides replicable and scalable Guangdong experience for the nationwide standardization of expert databases.

Firstly, it provides guidance for the governance of existing data. The referencing of the data specification offers a unified benchmark for cleaning, transforming, and integrating existing expert data, enabling originally fragmented and heterogeneous expert information to be governed according to consistent standards.

Secondly, it provides standardized operational guidelines for business processes. The value of the data specification lies not only in defining data standards but also in guiding practical business operations. From field definitions and data formats to validation logic, the data specification provides standardized guidance for the entire workflow of expert database management.

Thirdly, it serves as technical support for system upgrades and cross-regional collaboration. In terms of system upgrades, the specification defines clear data interface standards and system functional requirements, providing a technical basis for the upgrading and transformation of expert database systems. This facilitates seamless connectivity and data sharing between systems across regions.

Fourthly, it offers data-driven decision support. The implementation of the expert database data specification provides reliable data support and a basis for decision-making in specialized governance within the bidding and tendering sector. By analyzing data from the expert database, prominent issues in expert database management can be promptly identified and corrected. Relevant authorities can optimize the structure of the expert database in a targeted manner, improve the allocation efficiency of expert resources, and provide a precise tool for the specialized governance of prominent issues in the bidding and tendering sector.

4 Core Scenarios: From High-Quality Data to AI-Powered Intelligent Applications

The high-quality dataset constructed based on the Guangdong case can robustly support the following AI application scenarios, significantly enhancing the intelligence level of expert management.

4.1 AI-Powered Intelligent Selection and Precise Matching

Traditional expert selection primarily relies on fuzzy matching based on professional categories. Leveraging the high-quality dataset, AI enables the following capabilities. Multidimensional precise profiling integrates hundreds of data fields, such as an expert's professional domain, project experience, research achievements, and the scale and complexity of past projects, to construct an in-depth expert competency model. Intelligent recommendation and matching allows AI models, by considering the technical features, innovation requirements, and potential risks of specific bidding projects, to go beyond simple keyword matching and accurately recommend the most suitable candidates from a vast pool of experts, significantly enhancing the quality and efficiency of evaluations. Cross-regional optimal scheduling, under unified data standards, enables AI to break geographical constraints, identify the most suitable experts nationwide, and provide optimal team formation and network scheduling suggestions for "remote and cross-regional evaluations", truly realizing the nationwide sharing of expert resources.

4.2 AI-Assisted Evaluation and Risk Warning

High-quality historical evaluation data is crucial for training AI supervision models. Evaluation behavior analysis allows AI to examine behavioral patterns of experts in historical evaluations, such as scoring deviations, evaluation duration, and focus of inquiries, aiding in the identification of abnormal evaluation conduct and providing leads for supervisory work. Real-time conflict of interest verification enables AI, by combining standardized data like "recusal entity information", to perform automated, real-time conflict of interest scans before expert selection and evaluation, effectively preventing integrity risks. Dynamic performance capability assessment utilizes the analysis of management information, including experts' annual evaluations, training performance, and complaint resolution records, allowing AI to dynamically assess experts' performance capabilities and status, thereby providing data-driven decision support for their continuing education and rating adjustments.

5 Supporting the Unified National Market: The Multiplier Effect of Standardized Data

Guangdong's establishment of the first provincial-level data standard for a comprehensive evaluation expert database has broken through industry bottlenecks, and its practical significance holds profound implications for building a unified national market.

On one hand, it breaks down data barriers and promotes the flow of factors. Data standards serve as "soft infrastructure". When the same or compatible data standards are adopted nationwide, the technical foundation is laid for the joint development and sharing of a national expert database, enabling expert resources to flow efficiently across the national market like commodities.

On the other hand, it enhances regulatory efficiency and fosters a fair environment. Standardized data makes cross-regional collaborative supervision possible. Higher-level regulatory authorities can conduct macro-level analysis and risk monitoring based on unified data metrics, creating a more transparent, fair, and predictable market environment.

Lastly, it drives industrial upgrading and cultivates new quality productive forces. Empowering AI applications with high-quality data will propel the bidding industry from traditional manual management toward digital and intelligent transformation, improving total factor productivity. This represents a concrete manifestation of developing new quality productive forces in the field of public resource transactions.

6 Conclusion and Outlook

The case of Guangdong Province developing a data specification for its comprehensive evaluation expert database compellingly demonstrates that prioritizing data governance before technological application and leading the construction of high-quality datasets through standardization is a wise approach to unleashing AI potential and serving national strategies. It provides a clear pathway for transitioning from "seeking experts manually" to "AI recommending experts", and from "local utilization" to "nationwide sharing".

Looking ahead, it is recommended to advance efforts in two key areas: First, promote the elevation of the standard to a national standard or industry consensus, encouraging more provinces to draw insights from Guangdong's experience to achieve nationwide "uniform standards" for data. Second, foster innovation in AI applications based on standardized data. While ensuring data security and compliance, open data interfaces to attract stakeholders from industry, academia, and research to jointly develop smarter and more efficient application tools, ultimately injecting powerful digital momentum into the construction of a unified national market.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Chinese Communist Party Central Committee & State Council. (2022). Guideline on Accelerating the Establishment of A National Unified Market.Xinhua News Agency.
2. State Council. (2025). Premier Li Qiang presides over the 16th thematic study session of the State Council. Xinhua News Agency.
3. GB/T 36344-2018. (2018). Information Technology-Evaluation Indicators for Data Quality. Beijing: Standardization Administration of China.
4. T/DGAG 034-2025. (2025). Data specification for Guangdong Provincial Comprehensive Bid Evaluation Expert Database. Guangdong Association of Engineering Consultants.
5. National Development and Reform Commission. (2024). Measures for the Administration Bid Evaluation Experts and Bid Evaluation Expert Databases (Order No. 26). Beijing: NDRC.
6. Guangdong Provincial Government Data Administration. (2025). Measures for the assessment and management of bid-evaluation expert performance in the Guangdong provincial comprehensive bid-evaluation expert database (No. Yuezhengshu [2025] 13). Guangdong Provincial Government Data Administration.
7. Guangdong Provincial Government Data Administration. (2025). Administration measures for labor remuneration of bid-evaluation experts in the Guangdong provincial comprehensive bid-evaluation expert database. Guangdong Provincial Government Data Administration.
8. Guangdong Provincial Government Data Administration. (2025). Eligibility criteria for admission to the Guangd ong provincial comprehensive bid-evaluation expert database. https://zfsg.gd.gov.cn/zwgk/wjk/content/post_474 9068.html

Biographies

1. **Jiong Zheng** Bachelor, the Deputy Secretary-General of the Guangdong Digital Government Affairs Association, Senior Engineer, Hong Kong Registered Senior E-Government Engineer. He previously served as the Deputy Editor-in-Chief of the "Guidelines for E-Government Project Cost Estimation" and participated in numerous government informatization research projects and standard development;
2. **Haowei Qiu** Master, works at the Guangdong Provincial Government Service and Data Affairs Center;
3. **Lihua Huang** Bachelor**,** the System Integration Project Management Engineer, works at the Guangdong Digital Government Affairs Association.

構建高質量專家庫數據集
——以廣東實踐引領AI賦能與統一大市場建設

鄭炯[1]，丘皓瑋[2]，黃禮華[1]

[1]廣東省數字政務協會，廣州，中國，510030

[2]廣東省政務服務數據事務中心，廣州，中國，510030

摘要：在全國加快建設統一大市場的背景下，公共資源交易領域的標準化與數字化是關鍵一環。評標評審專家庫作爲招標投標活動的核心要素，其數據質量直接關係到資源配置的效率和公平。本文以廣東省綜合評標評審專家庫數據規範的創新實踐爲案例，深入探討了如何通過系統性數據治理，構建高質量專家數據集，爲AI智能抽取、評審輔助等應用場景的支撐，從而爲構建高標準、高效率的全國統一大市場提供堅實基礎。

關鍵詞：專家庫；數據規範；高質量數據集；人工智能

1. 鄭炯，學士，廣東省數字政務協會副祕書長，高級工程師/香港註冊高級電子政務工程師，曾擔任《電子政務工程造價指導書》副主編、參與多項政務信息化課題研究和標準編制；
2. 丘皓瑋，碩士，任職於廣東省政務服務數據事務中心；
3. 黃禮華，學士，系統集成項目管理工程師，任職於廣東省數字政務協會。