Article

Research on the Combination of Intelligent Management of Tax Data and Anti-Fraud Technology

Qifeng Hu [1,*]

[1] Independent Researcher, Springfield, Illinois, 62704, USA
* Correspondence: Qifeng Hu, Independent Researcher, Springfield, Illinois, 62704, USA

Abstract: With the continuous update and evolution of tax-related data, intelligent tax data management supported by big data-driven anti-fraud technologies has gradually become the core pathway for improving the efficiency, accuracy, and transparency of modern tax administration. Based on practical work scenarios, this study first outlines the theoretical framework of intelligent management, including data governance architecture, algorithmic decision-making mechanisms, and the functional logic of anti-fraud systems. It then systematically examines several prominent challenges currently faced in tax data management: inconsistent data standards across systems and departments, fragmented or interrupted process collaboration, lagging response of risk identification models, insufficient utilization of dynamic monitoring indicators, and relatively low interoperability among platform components. To address these issues, this paper proposes a set of targeted optimization strategies. These include establishing unified and fine-grained data standards to ensure semantic consistency, promoting cross-departmental collaboration through process re-engineering and automated workflow integration, and enhancing the responsiveness of risk detection models through dynamic model deployment, continuous training, and adaptive feedback mechanisms. Furthermore, the study highlights the importance of building an integrated tax governance platform that enables seamless data circulation, real-time communication across systems, and comprehensive risk visualization. The proposed solutions aim to provide actionable technical guidance for tax risk control, strengthen precision identification of abnormal behaviors, and enhance the capability of tax departments to detect, prevent, and respond to emerging fraud patterns in a timely manner. Ultimately, the research contributes to the modernization of tax administration and supports the development of a data-driven, intelligent, and resilient tax governance system.

Keywords: intelligent tax data; anti-fraud technology; integration issues; management mechanism; system integration

Published: 13 December 2025

1. Introduction

With the increasing digitization of tax management, the demand for timeliness, reliability, and confidentiality of data has become a key factor in measuring system efficiency. After the modernization of a series of related business procedures such as tax declaration, 1099 declaration, and corporate financial processing, the traditional method of relying on fixed audit rules for fraud risk assessment can no longer meet the diverse needs of fraud identification. At present, the US Internal Revenue Service and state finance bureaus need to further empower modern tax management through intelligent technology in areas such as data standardization and access, cross departmental workflow, and model feedback mechanisms. This article explores the integration of intelligent management and fraud prevention applications from both a technical and process perspective. It proposes a comprehensive solution that integrates data standards, process coordination, model linkage, and platform integration to promote the intelligent transformation of tax information exchange management and provide sustainable technical support for risk control mechanisms.

2. The Basic Logic of Intelligent Tax Data and Anti Fraud Technology

2.1. Technical Foundation of Intelligent Management

The intelligent management of tax data relies on a complete set of technology systems with information technology as the core, which mainly includes four levels: data collection automation, data fusion standardization, model algorithm intelligence, and platform operation integration [1]. At present, the integration capability of multi-source heterogeneous data has become a key starting point for system construction. Through API interface technology and ETL process automation tools, the system can uniformly capture and structurally integrate data sources from IRS electronic declaration systems, Form 1099 platforms, enterprise ERP accounting systems, and other sources. On this basis, with the help of feature engineering and data cleaning algorithms, a high-quality dataset that can be used for risk identification can be constructed. Subsequently, through intelligent algorithm models such as machine learning, graph computing, and correlation analysis, the system can identify suspicious patterns and generate preliminary risk warnings from large-scale data [2]. In addition, to ensure the continuous iteration capability of the model and real-time data response, it is necessary to build a support platform with distributed computing and automatic scheduling functions, so as to achieve the organic unity of dynamic controllability, data exchange efficiency, and intelligent decision support functions for tax data management.

2.2. Logical Framework of Anti Fraud Technology

The core logic of tax anti-fraud technology lies in relying on data-driven processes to identify, analyze, and respond to risks. The entire technical system can be divided into five main stages: data perception, feature extraction, risk modeling, warning triggering, and coordinated disposal. In the data perception stage, the system achieves real-time capture of business behavior by accessing multi-source data; Feature extraction focuses on transforming raw data into variables with recognition capabilities, including behavior frequency, transaction patterns, anomaly indicators, etc. Entering the modeling phase, the system utilizes clustering, graph algorithms, temporal analysis, and other methods to structurally model enterprise behavior and identify potential fraud risks [3]. The warning process filters high-risk behaviors and pushes warning information through threshold setting and model scoring; Ultimately, the linkage processing module connects the rule engine with the manual review mechanism to achieve policy response and model feedback updates. As shown in Figure 1, the five-segment logical path highlights the integrated characteristics of process collaboration, response loop, and model optimization, providing clear support for building an efficient and practical tax anti-fraud system.

Figure 1. Five link logical framework diagram of tax anti-fraud technology.

3. The Combination of Intelligent Management of Tax Data and Anti Fraud Technology

3.1. Inconsistent Data Standards

In the construction of intelligent tax systems, information data from different sources often have differences in format, structure, and naming, which become important reasons for slow data integration speed and poor data modeling in the later stage. Due to different sources, key attributes such as 1099 form data, personal tax return, and corporate tax return have different settings and numbering rules for different fields such as dates, amounts, NAICS industry codes, etc., which can lead to issues such as incompatibility, missing fields, and unclear meanings in data integration. In addition, some data management systems use local private file formats that cannot adapt to standardized data interfaces, making it impossible for these systems to be directly integrated in modeling tools or connected to anti-fraud systems, resulting in information silos in data processing [4].

3.2. There Are Breakpoints in the Management Process

To leverage the integration of tax data management and anti-fraud technology, it is necessary to build an information pathway that runs through the entire process. However, in actual operation, the interfaces between each department or system will always face structural breakpoints. On one hand, the process of data collection and analysis may be completed by the IT department or a separate data team, while risk response and subsequent control are respectively completed by the IRS's enforcement or fraud investigation department or the state taxation administration's enforcement unit. These departments have different goals, execution, and understanding of technology, and lack a standardized rule for process interoperability. On the other hand, the model recognition results are difficult to map to the business processing system in a timely manner. Due to the lack of a unified scheduling platform and standardized interface format, push delays or result loss occur, which reduces the effectiveness of risk response.

3.3. Lack of Responsiveness in Model Deployment

In the intelligent management system of tax data, the deployment efficiency of models directly affects the real-time and practicality of anti-fraud identification. However, in current practice, models often go through multiple layers of approval and manual operations from development to launch, with low update frequency and inability to respond in a timely manner to changes in new types of fraudulent behavior [5]. Some models are still embedded in the system in the form of static files, lacking automatic scheduling and version management functions, and unable to adjust discrimination rules based on real-time data, resulting in recognition delays. Some platforms have not established a unified model deployment framework, and different business systems need to interface with model services separately, resulting in complex interfaces and long deployment cycles, which affects the collaborative effect of cross system calls.

3.4. Insufficient System Platform Collaboration

The collaborative promotion of tax data management and anti-fraud mechanisms require multiple system platforms to achieve data sharing, functional interoperability, and task linkage. However, in actual operation, various business systems are often built by different vendors or development teams, lacking a unified technical architecture and interface standards, resulting in poor integration between systems and low information transmission efficiency. Some tax units have independent operation of business management platforms, data processing platforms, and risk analysis platforms, which not only result in lagging data synchronization, but also have issues with conflicting permissions and duplicated processes when calling functions.

4. The Combination Strategy of Intelligent Management of Tax Data and Anti Fraud Technology

4.1. Building a Unified Data Standardization System

The complexity of tax data is not only reflected in its diverse sources, but also in structural differences and semantic conflicts. Different business systems have inconsistencies in data item naming, field format, time dimension, industry coding, and other aspects, greatly limiting the efficiency of data sharing and invocation in anti-fraud models. Building a unified data specification system within the framework of intelligent management is a fundamental condition for bridging data fusion and model response.

Standardization work needs to be carried out synchronously from three levels: first, at the underlying structure, clarify the core field definitions shared by all business systems, unify field types, lengths, and naming rules; Secondly, in terms of exchange mechanisms, establish interface standards for cross platform data transmission, including data formats (such as JSON/XML), call frequency, verification mechanisms, etc.; Thirdly, at the semantic level, align the concepts of similar fields from different sources to avoid data bias caused by semantic ambiguity such as "amount", "tax amount", and "transaction amount". In addition, it is necessary to establish a data standard management platform to perform version control and permission management for standard updates, in order to avoid arbitrary changes in data definitions between different systems.

As shown in Table 1, the core building elements and implementation strategies for a unified data standardization system are:

Table 1. Construction Elements and Implementation Strategies of Unified Tax Data Standardization System.

Module category	Main content	Implementation strategy
Field Structure Specification	Unified field naming, format, and type	Establish a field dictionary and master data table for various systems to call
Data Exchange Protocol	Interface standards, call formats, authentication methods	Develop a unified API document and enforce standard interfaces
Conceptual semantic alignment	Same meaning for similar fields	Introduce a business semantic tagging system for field meaning mapping
Version control mechanism	Data standard maintenance and update process management	Set standard version release cycle, limit change process and review permissions
Data quality verification	Consistency, completeness, and timeliness verification mechanism	Embed an automatic verification module to record quality logs and exception reports

Building a unified regulatory system is not a one-time project, but should become a part of data lifecycle management and be incorporated into the daily mechanism of data governance. By using standardized methods to reduce structural complexity, the quality of model training, recognition accuracy, and platform collaboration efficiency can be significantly improved.

4.2. Optimize Cross Departmental Process Collaboration Mechanism

The key to intelligent management of tax data lies in achieving smooth flow and efficient cooperation between various links in the data chain. However, in actual operation, data collection, cleaning, modeling, warning, feedback and other links are scattered in different departments or systems, resulting in unclear division of responsibilities, chaotic interface permissions, and inconsistent process responses, which directly affect the efficiency and closed-loop capability of anti-fraud work.

To optimize the cross departmental collaboration mechanism, the primary task is to build a unified process coordination platform, clarify the responsibility attribution, triggering conditions, and output requirements of each link, and ensure the continuity of data flow and task response time at different stages. The efficiency of collaborative processes can be simplified into the following calculation formula:

$$E = \frac{D}{T_1+T_2+T_3+\cdots+T_n} \quad (1)$$

Among them, E represents the overall process efficiency, D is the system goal (such as model update cycle, risk response window), and T_1 to T_n are the task processing time of each key link. The optimization objective is to minimize the total duration of the process while maintaining the stability of the objective, in order to enhance the immediacy of risk response.

At the same time, a cross institutional data coordination center should be established to distribute model recognition results to business units such as the IRS audit department, criminal investigation department, and data monitoring team through an automated process platform, reducing human intervention and improving risk linkage efficiency. The scheduling rules should have dynamic adjustment functions and support automatic diversion based on dimensions such as risk level, industry type, or geographical distribution.

4.3. Strengthen the Dynamic Linkage of Model Deployment

In intelligent tax management, the design of anti-fraud models is not a one-time integrated task, but rather requires the ability to respond in real-time and dynamically correlate. However, in actual operation, it can usually only be completed through manual or offline input methods, resulting in inconsistency with real-time data systems, leading to model recognition lag and poor efficiency. Therefore, it is necessary to develop a dynamic installation technique for models in combat situations, where the model can be dynamically associated with changing data, business commands, and platform assignments in real time.

We should design an end-to-end collaboration mechanism that covers three dimensions: firstly, to achieve process intelligence. By using the CI/CD toolchain, the entire process of training, testing, and validation of the model can be automatically completed according to predetermined steps; The second is automatic strategy adjustment, which triggers the redeployment of the model and automatic adjustment of strategies based on actual situations when risks change, data intensity or model performance indicators need to be adjusted; The third is to build a feedback loop. The identification results generated by the model are automatically fed back to the central training stage, allowing for timely fine-tuning and maintenance of version control. The efficiency of model linkage can be measured by the following indicators:

$$L = \frac{P \times R}{D+\Delta} \quad (2)$$

Among them, L is the linkage index, P represents the concurrent deployment capability of the model, R represents the response rate, D is the data update delay, and Δ is the deployment trigger delay. This formula is used to measure the deployment response level of a system in high concurrency and complex scenarios. The higher the linkage index, the timelier the system deployment, the more compact the feedback, and the more stable the risk identification ability.

In addition, a centralized control model library and model calling platform should be built to centrally manage various risk identification models, versions, parameters, usage history, and other metadata, avoiding the unacceptable difficulties caused by integrating multiple separate systems. It should also have functions such as visitor permission management, collaborative calling, and status tracking to ensure consistent use and joint response of various models in the same business application.

Deploying dynamic linkage not only requires technical support, but also requires the reconstruction of the management system. It is necessary to incorporate the lifecycle of the model into the dynamic management of the tax governance system, and to achieve real-time linkage and closed-loop development of strategies, algorithms, and data.

4.4. Promote the Integrated Operation of the System Platform

In the context of the deep integration of intelligent management of tax data and anti-fraud applications, the integration level of the system platform directly affects the efficiency of data circulation and the stability of model operation. Currently, most institutions still face the problem of scattered deployment and independent operation of business platforms, data warehouses, model engines, and risk control modules, leading to interface duplication, task scheduling conflicts, and lagging data synchronization. To improve overall collaboration capabilities, it is necessary to promote the unified architecture upgrade of the existing IRS information system, and achieve efficient connectivity and centralized management between the tax data platform, fraud identification engine, and risk response system.

The core goal of platform integration operation should be "sharing, collaboration, and controllability": sharing refers to unifying data interface standards, ensuring that data between systems can be mutually recognized, fields can be mapped, and semantics can be decoded; Collaboration refers to building a unified task scheduling center to allocate the execution sequence and frequency of functional modules such as model calls, warning push, and data feedback; Controllable refers to strengthening system permission management, monitoring operation logs and abnormal alarms, achieving operational transparency and traceable results. To quantify the effectiveness of platform integration, the following calculation formula can be introduced:

$$C = \frac{M_s \times D_r}{I_o + T_c} \tag{3}$$

Among them, C represents the integration efficiency index, M_s represents the integration degree of model services, D_r represents the coverage of data calls, I_o represents the redundancy of system interfaces, and T_c represents the time consumption of cross platform calls. The larger the value, the higher the system integration efficiency, the more optimized the resource allocation, and the closer the operating efficiency is to the target state.

In terms of platform design, a microservice architecture should be adopted to decouple core functions into independent modules and achieve high availability and horizontal scalability through containerization and service grids. Simultaneously introducing a unified identity authentication system and permission management mechanism to avoid duplicate login and permission conflicts across multiple platforms. At the business application layer, the front-end interaction interface style should be unified, integrating model push, result query, and risk feedback channels to improve the operational efficiency of frontline users.

The integrated operation of the platform is not the integration of a single module, but the reconstruction of the entire system logic. The supporting environment of the intelligent tax anti-fraud system that breaks through the platform boundary keeps the entire platform unified in logic, data and service layers.

5. Conclusion

The integration of intelligent management of tax data and anti-fraud technology not only helps improve the enforcement of federal tax laws, but is also an essential solution for precise tax supervision and effective management. This article presents a solution centered around four key issues: data standardization, workflow coordination, model deployment, and system integration. The solution is characterized by rule-based guidance, linked by joint response, and centered around platform integration. Only by promoting technological collaboration and mechanism connectivity between various platforms can we build an efficient, stable, and sustainable anti-fraud support system. The tax department needs to enhance the construction of intelligent models in future development, optimize the exchange mode of data among various departments, improve integration, and shift anti-fraud capabilities from local response to full chain linkage. This study aims to provide practical ideas and technical references for the construction of intelligent systems and the design of risk identification mechanisms in the field of taxation, further promoting the structural reshaping and intelligent evolution process of tax information systems.

References

1. B. Fan, L. Sun, D. Tan, and M. Pan, "Optimal Selection Technology of Business Data Resources for Multi-Value Chain Data Space-Optimizing Future Data Management Methods," *Electronics (2079-9292)*, vol. 13, no. 23, 2024. doi: 10.3390/electronics13234690
2. C. Cheng, B. Li, and D. Chen, "A quality evaluation method for intelligent traffic management system based on business data flow," *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 4, pp. 6193-6208, 2023. doi: 10.3233/jifs-230182
3. O. E. Akinbowale, H. E. Klingelhöfer, and M. F. Zerihun, "Investigating the level of effectiveness of the anti-fraud technologies employed by the South African banking industry for cyberfraud mitigation," *Journal of Financial Crime*, vol. 31, no. 1, pp. 201-225, 2024. doi: 10.1108/jfc-02-2023-0025
4. J. A. Summers, N. Wilson, T. Blakely, and F. Sigglekow, "Disease-Related Loss to Government Funding: Longitudinal Analysis of Individual-Level Health and Tax Data for an Entire Country," *Value in health*, vol. 26, no. 2, pp. 170-175, 2023. doi: 10.1016/j.jval.2022.08.007
5. V. S. Verykios, E. C. Stavropoulos, V. Zorkadis, G. Katsikatsos, and E. Sakkopoulos, "Sensitive data hiding in financial anti-fraud process," *International Journal of Electronic Governance*, vol. 14, no. 1-2, pp. 7-27, 2022.