Current Status and Normative Optimization of the Data Intellectual Property Registration Framework

Haojie Dong
China Jiliang University, Hangzhou, China

Abstract: To further promote data circulation and transactions, China has innovatively established a data intellectual property registration system, currently implemented through pilot programs across multiple regions where it has demonstrated tangible outcomes and institutional value; however, existing practices reveal persisting deficiencies in defining registrable objects, clarifying entitlement attribution, and setting protection durations. Consequently, integrating regional experiences, the system requires uniform standards for nationwide consistency in registrable objects, explicit allocation of entitlement exclusively to data processors (excluding originators as rights holders), and the establishment of rational protection timelines through refined regulations—only through these measures can the system effectively unleash its potential to energize new quality productive forces.

Keywords: Digital; Data Property Rights; Data Intellectual Property; Registration; Intellectual Property Registration System

1. Empirical Analysis of Local Data Intellectual Property Registration Practices

1.1 Scope of Data Intellectual Property Registration

Current provincial and municipal Data Regulations universally establish that data processors hold statutory rights to autonomously utilize, dispose of, and derive benefits from lawfully generated data products. Theoretically, regarding the registrable objects of data intellectual property, scholars contend that the eligible subject matter should be "publicly accessible data collections." However, whether all publicly available collections warrant legal protection remains unclarified.[1]The registration object involves the specific issue of the object that can or should be registered, which is a fundamental issue for exploring the carrier of data rights and aims to resolve the controversial focus on the object of data intellectual property. Theoretically, regarding the registration object of data intellectual property, some scholars believe that the object of data intellectual property should be an "open data set", but there is no conclusive conclusion on whether all open data sets should be protected by law.[2] Some scholars argue that the registration object of data intellectual property is a "data product formed by deeply processing massive chaotic and unordered raw data using algorithms and analysis models", which limits the object of data intellectual property to data products.[3]

This legislative trend acknowledges enterprises' foundational claims over specific data assets at a macro level. Notably, local regulations and the data intellectual property registration system jointly construct a multi-tiered normative framework. Empirical analysis of registration measures from 21 provinces/municipalities (including Yangtze River Delta and Pearl River Delta regions) reveals five core requisites for data intellectual property registration:

Lawfulness constitutes the fundamental threshold for system access. Local regulations explicitly require data products to originate from "legally collected" and "legitimately processed" sources. The term "legally" carries dual meanings: compliance with higher-level legal frameworks (e.g., Civil Code, Personal Information Protection Law) and adherence to special provisions in local normative documents.

Direct protection of raw data is universally excluded. Registrable subject matter must be structured datasets processed by algorithms, yet legislation lacks quantifiable thresholds for technical processing. This ambiguity has sparked practical disputes—e.g., a Shenzhen big data company's application was rejected for using basic cleansing algorithms, while a Shanghai counterpart secured registration with identical techniques. Despite implementation variances, the "Minimum Processing Principle" has gained consensus, creating an institutional barrier

against propertizing raw data.

This requirement substantially complements data form criteria. Processed datasets must demonstrate market viability or practical utility, compelling registration authorities to establish professional evaluation systems. Zhejiang Province has piloted a joint assessment panel (data scientists, economists, jurists) implementing tripartite evaluation (technical value, transaction potential, social benefit). However, this mechanism fails to fundamentally resolve standardization challenges.

The intellectual contribution attribute—intrinsic to IP rights—is embedded in the system. Unlike traditional IP, data IP requires neither copyright-level originality nor patent-grade novelty. Jiangsu Province's draft "creative labor" standard has provoked academic debate for blurring IP protection boundaries and risking institutional distortion.

Beijing uniquely mandates "non-public status" as a prerequisite but faces dual interpretative ambiguity: whether it refers to the entire dataset being undisclosed or individual data points remaining non-public. This vagueness creates operational challenges. Most provinces omit this requirement, reflecting deliberate neutrality toward the theoretical controversy regarding data value's correlation with public disclosure status.

1.2 Data Intellectual Property Registration Procedure

The registration procedure serves as the technical core of local data intellectual property rule systems. Its institutional design involves three key regulatory dimensions: application, content specifications, and examination. Despite variations in implementation details across provinces and cities, the institutional core demonstrates significant convergence.

While different provinces and cities emphasize different aspects in their procedural chapters and implementation rules, the registration application mechanism shares three universal characteristics. Regarding applicant eligibility, all regions confirm the foundational status of data processors. Units or individuals that lawfully collect and process data naturally possess data rights and enjoy registration application eligibility based on their substantive labor input in data processing activities. As for application methods, the system design reflects flexibility: supporting applicants to file independently or entrust professional agencies to act on their behalf; for data products jointly created by multiple parties, all participants must apply jointly unless otherwise agreed upon. It is particularly noteworthy that Zhejiang Province requires completion of a data attestation procedure before registration application can be made. This may be achieved through notarization by a public notary office or via technological means such as blockchain. This regulation aims to enhance the verifiability and traceability of registered data. For example, a Hangzhou-based biomedical company shortened its examination cycle by 37% through blockchain attestation. While other provinces do not mandate attestation, pilot programs like Shanghai have begun implementing incentive mechanisms for attestation.

When stipulating data intellectual property registration, jurisdictions universally employ a field description approach to specify the registered data, differing only in the level of detail and specific requirements.[4] The registered content must encompass five core elements: dataset name; industry sector and application scenarios; data source and formation time; structural scale, algorithmic rules and update frequency; and attestation/notarization status. These elements collectively serve as the technical vehicle for publicizing rights, whose core function is to disclose rights boundaries and protection scope to the public. Given the intangible nature of data and the intertwining of diverse interests, the registration system must simultaneously achieve dual safeguards: specific identification of the rights object and legitimacy verification of the rights source. Specific identification requires locking down the protected object through description of key information. Foremost is standardized naming of the dataset, which may be based on data sources, industry fields or application scenarios (e.g., "Yangtze River Delta New Energy Vehicle Supply Chain Map"), accompanied by a unique identification number. Even when data sources are similar, such naming conventions clearly differentiate distinct data products arising from differing application scenarios or analytical methods. Algorithmic rules and structural features constitute the core technical elements of registration. Algorithms embody the intellectual labor input of data processors, while data structure reflects the organizational form of data. Their combination objectively presents the technical substance of the data product and thus

should be required registration items. Legitimate verification of data sources is a baseline requirement of the registration system. Applicants must submit documentary proof of lawful data acquisition and processing, while publishing a declaration of legitimacy in the registration record. This serves both as the foundation for legitimacy review of the rights source and complies with the regulatory requirements of Article 13 of the Personal Information Protection Law and Article 32 of the Data Security Law. Given the mutable nature of data, attestation/notarization becomes indispensable evidentiary support for subsequent rights protection.

Registration examination involves issues of review standards. Current local legislation predominantly adopts formality examination for efficiency considerations, meaning it reviews only whether application materials meet prescribed requirements and comply with registration procedures, without verifying the authenticity of content matters. [5]The widespread adoption of formality examination across regions is largely due to efficiency concerns. During the pilot phase of data intellectual property registration, not only is data intellectual property itself not yet explicitly recognized by law, but consensus is also lacking on registration conditions, recognition standards, relevant procedures, and responsible entities. Formality examination therefore reduces procedural complexity, avoids imposing excessive costs on relevant departments, and encourages more enterprises to pursue data intellectual property registration. However, while formality examination increases registration efficiency, it carries inherent risks regarding the authenticity of registered content and submitted materials. To address this, regional registration management regulations have implemented supplementary measures. For instance, Shenzhen employs a pre-regulatory approach requiring substantive examination by third-party service institutions for initial data property registrations; applicants must submit materials reviewed for authenticity and legality by these third-party institutions. Current practice shows relatively few data intellectual property registration activities, and instances of fraudulent registration through submission of false materials have not yet been identified. Thus, whether formality examination is appropriate for data rights registration awaits further practical validation through expanded case studies.

1.3 Legal Effect of Data Intellectual Property Registration

Absent a national statutory framework establishing independent data property rights, local registration practices cannot confer novel intellectual property rights upon data holders. This fundamentally differs from patent or trademark systems, as local registration lacks right-creating effect and generates no exclusive legal consequences. The current regime, derived from policy instruments like the Twenty Data Measures, serves exclusively as a public attestation mechanism rather than a right-granting tool at both central and local levels. Consequently, without national legislation creating new data property rights, local data IP registration carries no enabling legal effect. This means registration neither alters the essential nature of rights nor substantively modifies data holders' control status and entitlements over data resources. Registration certificates merely provide administrative confirmation of pre-existing factual conditions and create no new intellectual property rights. The system essentially functions as a public certification service provided by administrative agencies amid legislative gaps, akin to real estate registration which offers public notice effect without creating new rights.Scholars recognizing only the certificatory function of registration maintain that all data property registration—not limited to data intellectual property—serves solely to evidence entitlement attribution and scope, aiming to reduce transaction costs in data rights transfers or exchanges, and to protect data rights while safeguarding transaction security.[6]

Though incapable of creating new property rights, data IP registration delivers critical evidentiary functions as an administratively administered public attestation procedure. Registration authorities establish a dual-track mechanism: registration databases publicly disclose rights attribution, boundaries, and attestation codes to unspecified parties, leveraging the public trust principle under the Civil Code; while registration certificates serve as foundational proof of ownership for market activities, providing transaction security endorsement in data exchanges and financing. During disputes, duly issued certificates qualify as official documentary evidence under the Civil Procedure Law, bearing superior evidentiary

weight in judicial proceedings.

2. Emerging Issues in Local Practices of Data Intellectual Property Registration

2.1 Definitional Challenges Regarding the Scope of Registrable Objects

Current local pilot practices have reached preliminary consensus on the scope of registrable objects for data intellectual property registration, yet significant theoretical divergences and practical controversies persist regarding specific protection requirements, primarily manifesting in two challenges: the assessment of innovativeness and the regulation of public disclosure. When the object of data intellectual property is limited to data sets or data products with practical value and the attribute of intellectual achievement, the issue of registration capacity must be considered to determine whether they can be registered. The so-called registration capacity refers to the "qualification" to connect with the publicity methods required by law, to have identifiability after registration, and to be independent.[7]

In terms of innovativeness assessment, local regulations generally require data products to demonstrate "creative labor," yet the inherent characteristics of data fundamentally differ from traditional intellectual property. Based on processing depth, data products can be categorized into two types: aggregated products undergo basic processing like anonymization and cleansing, their level of innovation being substantially lower than the originality threshold required for copyright or patents; derivative products involve deep algorithmic transformation but typically still fail to meet patent novelty standards. This dilemma leads to irreconcilable conflicts in protection standards: if intellectual property innovation requirements are strictly applied, commercially valuable aggregated products would be excluded from protection; if innovation thresholds are lowered, existing intellectual property systems risk destabilization. Consequently, critical questions demand resolution: must data intellectual property contain creative elements? If so, what tiered innovation standards should be established?

Regarding disclosure regulation, competing interests create governance paradoxes. Based on their public-domain nature, disclosed data should in principle be freely accessible. However, Beijing's pilot program imposes a "non-disclosed" registration threshold, while other regions set no such restrictions, reflecting fundamental philosophical conflicts. The core issue lies in whether granting exclusive rights over disclosed data—after creators have profited from its disclosure — violates patent law's foundational quid pro quo principle (disclosure for protection) and obstructs data mobility. As a landmark case illustrates, courts acknowledged the commercial value of a mapping service provider's disclosed geospatial data but denied the existence of exclusive rights against web scraping. The data intellectual property registration system urgently must resolve: can exclusive rights attach to disclosed datasets? How can investor returns be reconciled with public data accessibility?

2.2 Determination Dilemma of Entitlement Attribution

The core controversy in data intellectual property ownership fundamentally concerns the distribution of interests between data processors and data originators. Judicial practice has affirmed the lawful rights of data processors over derivative data products, yet the legal positioning of individuals within enterprise data rights frameworks remains unresolved, primarily due to two interconnected tensions:

From the perspective of data formation, users substantively participate in value creation. Data production involves two interdependent stages: first, users actively provide or passively generate raw data through online activities, forming the "data raw materials" for processing; subsequently, enterprises extract value through algorithmic analysis, transforming data into commercial assets. This creates a symbiotic relationship where daily user behaviors inherently serve a data-supplying function, while platforms act as resource integrators. When users meaningfully contribute to the value of final data products, whether they should share corresponding rights emerges as a critical legal question demanding resolution.

Furthermore, the dual nature of personal data rights intensifies ownership disputes. On one hand, personal data embodies human dignity, with Article 1034 of the Civil Code explicitly classifying health information, location tracks, and similar data as sensitive personal information warranting enhanced protection.

On the other hand, the proprietary value of personal data has been empirically validated. This duality necessitates institutional designs that simultaneously: prevent enterprises from exploiting personal data for profit without constraint, yet acknowledge the substantive contributions of originators to raw data formation. Consequently, individuals occupy a unique position within the data rights structure—neither meriting complete rights over final data products nor having their foundational contributions disregarded. When enterprise-processed data products incorporate personal data elements, balanced mechanisms must reconcile originators' personality rights with processors' proprietary interests.

2.3 Protection Period Setting Dilemma

The establishment of protection periods for data intellectual property registration carries significant systemic value, with its core function being to balance the interests of data rights holders against the public interest, thereby preventing perpetual monopolization of data resources. However, this institutional framework faces dual practical challenges:

First, disparities in data lifecycles create duration adaptability dilemmas. The value-decay cycles of distinct data types vary substantially—for instance, medical records may retain analytical value for years, whereas e-commerce transaction data often requires updates within months. Adopting the EU Database Directive's uniform 15-year protection model proves increasingly incompatible with modern data iteration speeds. Such duration constitutes overprotection for short-lived data, hindering public domain reuse, yet under-protection for long-lived data, discouraging sustained investment. Data's temporality necessitates graded protection mechanisms allowing high-frequency data to enter public circulation expediently.

Second, dynamic updating introduces protection period abuse risks. Continuous dataset updating raises concerns about prolonged protection terms. Rights holders might employ automated processes to make minor incremental additions (e.g., monthly 0.5% content updates), strategically claiming renewed protection periods. This parallels copyright's "de facto term extension" dilemma, fundamentally contravening intellectual property's temporal limitation principles. Jurisprudence already reflects this tension: A navigation service provider's claim for perpetual rights based on continuous traffic data updates was rejected under judicial "substantial change" standards. This underscores the need for update thresholds (e.g., 30% content replacement) as objective benchmarks for recomputing protection periods, preventing technical manipulations that undermine term-limitation regimes.

3. Refining the Data Intellectual Property Registration Framework

3.1 Establishing Unified Registrable Objects

Data intellectual property registrable objects refer to data products formed through lawful collection and processing, which exhibit a defined scale and commercial value, are subject to management measures, and embody attributes of intellectual achievements. Scholars Huang Hui and Yin Pengxu advocate that, considering the trial-and-error costs of the registration system and registration efficiency, data registration should be established as a new type of intellectual property registration.[8]Scholars Meng Qixun, Cheng Weijia, and Dai Yun argue that from the perspective of legal relations, data intellectual property registration takes specific data sets as the object, data processors as the main body, and limited exclusive rights as the content of rights.[9]This category specifically includes aggregated data products and derivative data products—that is, data collections that have undergone substantive processing — while expressly excluding raw data.

Registrable objects must cumulatively satisfy six requirements: legality requires data to be lawfully generated under China's regulatory regime, ensuring lawful sourcing and non-conflict with prior rights as the foundational prerequisite; innovative character recognizes that while data typically falls below traditional IP creativity thresholds, enterprises' intellectual labor in algorithmic processing (selection, classification, organization, mining, etc.) inherently confers intellectual achievement attributes sufficient for novel IP qualification; disclosable nature demands the object's capacity for public disclosure as a core registration precondition; manageability necessitates rights holders demonstrating both subjective management intent and objectively verifiable control measures externally perceptible; commercial value is established through judicial

precedents universally recognizing economic value derived from processors' creative investments in data products; and quantitative threshold requires registrable collections to meet minimum volume baselines — individual data points being insufficient for big-data analytics—with regional practices empirically validating the 30,000-unit benchmark as industry-compatible scale minimum.

3.2 Clarify Entitlement Attribution in Data Intellectual Property Registration

The rights holder in data intellectual property registration is the data processor, typically an enterprise. Data originators are not rights holders under this system, nor do they share rights with processors. Local pilot systems recognize two categories: (1) data compilers, acquirers, or custodians; (2) data processors or developers, including individuals, legal persons, and unincorporated organizations. For collaborative processing, all parties apply jointly; for commissioned processing, clients or both parties apply per contractual terms. Applicants may file independently or authorize agents—practices mirroring traditional IP registration that should be retained. However, institutional safeguards must protect data originators' interests in personal data.

First, vesting data IP rights solely with processors finds theoretical and practical justification. The labor entitlement theory provides the core legal basis: when processors transform raw data into commercially viable products through specialized labor, they generate qualitative value shifts—from low-value raw states to market-valued intellectual assets—warranting proprietary recognition. Utilitarian analysis reveals deeper societal benefits: granting processors exclusive rights incentivizes data production, fuels efficient resource utilization, and stimulates digital economic growth. The Coase Theorem's market efficiency principle further confirms that centralizing rights with end-to-end processors avoids transaction costs from rights fragmentation, optimizing resource allocation.

Second, individuals fundamentally lack qualification as data IP rights holders. Although their role in raw data generation merits attention, granting them IP status faces structural barriers: while individual data fragments hold micro-value, processors' algorithmically synthesized aggregate value transcends mere summation. Our impact assessment reveals a dual dilemma upon individual empowerment: micro-level bargaining asymmetry turns user agreements into de facto rights transfers; macro-level repetitive authorization requirements would escalate compliance costs. Such overprotection ultimately stifles data market vitality, contradicting digital economy development objectives.

3.3 Establish Appropriately Calibrated Registration Protection Periods

The establishment of protection periods for data intellectual property registration must balance its unique temporal characteristics with public interests. Compared to traditional intellectual property objects such as real estate and patents, data undergoes faster value depreciation and possesses inherent public attributes. Excessively long protection periods hinder the release of data's societal value. A five-year baseline protection period is recommended, after which rights automatically expire (without renewal mechanisms), and relevant data immediately enters the public domain for societal development and utilization.[10]

To address practical conflicts arising from data's variable timeliness—where certain data loses value within months while rights holders continue monopolizing idle resources—a tiered annual fee system should be introduced: Rights holders pay progressively increasing annual fees. This economic disincentive compels enterprises to continuously evaluate data value, voluntarily relinquishing exclusive rights to low-efficiency data, thereby facilitating data circulation. This design prevents idle data from obstructing circulation while ensuring continuous protection for high-value data.[11]

4. Conclusion

The dynamic value of data as a core element of the digital economy lacks corresponding institutional support. To address this challenge, China has pioneered an innovative data intellectual property registration mechanism, implementing differentiated pilot programs at local levels. Pilot regions have achieved preliminary institutional outcomes in defining rights holders, delineating registrable objects, and establishing verification pathways. However, constrained by insufficient theoretical foundations and the profound systemic impact of such institutions, a statutory framework for new-

type data IP registration remains unestablished at the national level. This directly results in inherent limitations to local registration effectiveness, failing to adequately support standardized data circulation and rights protection. Consequently, issues emerge including inconsistent registrable objects, ambiguous ownership attribution, and undefined protection periods. China urgently needs to construct a tiered legal framework for data IP, yet its design confronts multiple jurisprudential dilemmas. Constrained by current practical depth and theoretical maturity, proposed solutions remain overly abstract, requiring further refinement in implementation rules. Data IP registration represents a critical direction for future data governance. With evolving academic theories and accumulated practical experience, China will progressively develop a distinctive data IP registration model.

References

[1] Xu Shi. Intellectual Property Paths for Enterprise Data Protection and Their Breakthroughs. East China Law Review, 2018(5):55-62.

[2] Liu Xin. Rationale and Approach for Legislation on Data Intellectual Property in the Big Data Era. Intellectual Property, 2023(11):1-16.

[3] Gao Yang. Theoretical Justification of Derivative Data as a New Type of Intellectual Property Object. Social Sciences, 2022(2):1-12.

[4] Liu Wei. On the Rights Allocation of Data Products. Peking University Law Journal, 2023, 35(6): 120-145.

[5] Liu Xiaochun, Du Tianxing. Confirmation Registration of Data Intellectual Property Rights in the Construction of Data Element Market. China's Foreign Trade, 2023(7):1-20.

[6] Cheng Xiao. On Data Property Registration. Law Review, 2023, 41(4):20-35

[7] Li Yongjun. On the Influence of "Registration Capacity" of Property Rights on the Validity System of Real Rights. Law and Business Research, 2021(6):1-15.

[8] Huang Hui, Yin Pengxu. Research on Data Sharing Issues from the Perspective of the Public Domain. Journal of East China University of Political Science and Law, 2023(6):63-76.

[9] Meng Qixun, Cheng Weijia, Dai Yun. Research on the Pilot Reform Path of China's Data Intellectual Property Registration System. Science & Technology Progress and Policy, 1-10.

[10] Tang Zhenyou. Institutional Logic and Perfection of Data Intellectual Property Registration. Intellectual Property, 2024(3):34-45.

[11] Lü Bingbin. Data Intellectual Property Registration: Trade Secret Model or Database Model. Intellectual Property, 2024(6):62-79.

