AI and Data Science Journal
ISSN: 3078-7394
DOI: https://doi.org/10.61784/adsj3012

LEGAL ATTRIBUTES OF COMMERCIAL DATA AND APPROACHES TO INTELLECTUAL PROPERTY PROTECTION

Zhaoxia Deng
Guangdong University of Foreign Studies, Guangzhou 510515, Guangdong, China.
Corresponding Email: Angelina_dzx@163.com

Abstract: As a core production factor in the digital economy, the legal attributes and protection pathways of commercial data have become a research hotspot in the field of legal studies. This paper begins by examining the legal nature of commercial data and its attributes under intellectual property rights, establishing the legitimacy foundation for intellectual property protection of commercial data through labor property theory, utilitarian theory, and transaction cost theory. Besides, it proposes a "limited rights + fair use" governance model to address current legal dilemmas in data rights allocation. The study further recommends approaches such as strengthening technical governance and collaborative regulation, promoting international rule harmonization, to balance innovation incentives with public interests. Ultimately, these measures aim to achieve efficient circulation and value realization of commercial data.
Keywords: Commercial data; Legal attributes; Intellectual property protection; Fair use; Technical governance

1 INTRODUCTION

In the era of digital economy, data, as a new type of production factor, has become the core driving force for social transformation and economic growth. Commercial data, due to its unique economic value and strategic significance, plays an irreplaceable role in industrial innovation, market competition, and social governance. However, as the value of data resources has increasingly emerged, issues such as vague legal attributes, disputes over rights ownership, and divergences in protection paths have gradually come to the fore. The traditional legal framework faces structural challenges in addressing data rights confirmation, circulation, and benefit distribution: on the one hand, characteristics such as the non-exclusivity and replicability of data make it difficult to fully fit into existing rights systems such as property rights or intellectual property rights; on the other hand, excessive protection may inhibit data sharing and technological innovation, while insufficient protection is prone to triggering "tragedy of the commons" or "free-riding" behaviors, leading to market failures. Against this backdrop, clarifying the legal attributes of commercial data and exploring an adaptive intellectual property protection approach has become an urgent proposition to be solved in legal research. Starting from the legal attributes of commercial data, this paper systematically demonstrates the justificatory foundation for its intellectual property protection by integrating theories of legal philosophy and law and economics. Aiming at the dilemmas of "over-generalization of rights" and "circulation blockages" in current empowerment practices, it proposes corresponding countermeasures and suggestions, with the aim of providing theoretical references for constructing a data governance framework that balances innovation incentives and public interests.

2 ANALYSIS OF THE LEGAL ATTRIBUTES OF COMMERCIAL DATA

As a new type of production factor in the digital economy era, commercial data exhibits unique attributes in physical, economic, and legal dimensions. At the physical level, data exists in the form of binary code and has three technical characteristics: non-exclusivity (the same data can be held by multiple subjects simultaneously), replicability (marginal cost approaches zero), and non-consumability (use does not lead to value depreciation). At the economic level, data resources demonstrate dynamic value (value-added effects through processing), application scenario dependency (value varies with the application environment), and compound rights and interests (interweaving interests of multiple subjects). At the legal level, there are normative challenges such as the virtualization of object form (lack of physical boundaries of traditional property rights objects), ambiguity in rights ownership (conflicts between the rights and interests of data producers, processors, and users), and hierarchical protection needs (differences in protection intensity between raw data and derivative data)[1].

There is a theoretical confrontation in current academic circles regarding whether commercial data can constitute the object of intellectual property rights. From the perspective of traditional theoretical frameworks, opponents put forward three main objections: first, the lack of an originality requirement, as raw data, as a record of objective facts, does not meet the originality standard of copyright law; second, the impossibility of exclusive control, as the technical characteristics of data circulation inherently conflict with the exclusivity of intellectual property rights; third, a misalignment in the protection system, arguing that existing systems such as trade secrets and anti-unfair competition laws are sufficient to regulate data rights and interests[2]. Proponents, however, argue that derivative data products, formed through algorithmic processing, contain original content, and their nature as intellectual achievements is homogeneous with the objects protected by intellectual property rights. Based on the practices of digital economic development, they emphasize three aspects of protectability for data products: derivative data reconstructed through algorithms possess the attributes of intellectual achievements (such as user behavior analysis models); databases formed

through processing massive data meet the requirements for compiled works (such as product information databases on e-commerce platforms); and commercial intelligence formed by specific data combinations conforms to the criteria for trade secrets (such as precision marketing databases)[3].

This paper argues that commercial data refers to a collection of derivative data with market value, formed through large-scale extraction and processing of raw data via intellectual labor, and filtered and intelligently pushed for sales through algorithmic technology. Raw data, lacking creative input and independent economic value, is not suitable for direct rights assignment; by contrast, derivative data products, through human intellectual activities, form new information configurations that meet the core requirements for intellectual property protection.

3 POSITIONING THE ATTRIBUTE OF INTELLECTUAL PROPERTY RIGHTS FOR COMMERCIAL DATA

Although derivative data can become the object of intellectual property protection, the intellectual property rights in commercial data still differ significantly from traditional intellectual property rights. This section discusses the differences mainly from three aspects: the object of rights, the content, and the effect of rights.

3.1 The Dynamic Nature of Rights Objects

The objects of traditional intellectual property rights exhibit distinct static characteristics. For example, the expressions protected by copyright law—works—must have a fixed form at the time of creation (such as literary, musical, or artistic works), and their rights boundaries are defined directly by the author's creative act. Technical solutions, as objects protected by patent law, must be disclosed in a specification and form definite claims, with the scope of rights limited to the authorized text and subject to strict restrictions on subsequent modifications. For trademarks, the distinctiveness of the mark must comply with legal requirements at the time of registration, and the scope of rights is limited to the approved mark and the goods or services for which it is registered.

In contrast to the objects of traditional intellectual property rights, commercial data, as the object of data intellectual property rights, demonstrates significant dynamic characteristics. First, data aggregation is fluid. Derivative data products process massive raw data in real time through algorithms, and their content evolves dynamically as data sources update and models optimize[4]. For example, a traffic flow prediction model must continuously incorporate real-time traffic data to maintain accuracy, causing the boundary of the rights object to change continuously. Second, the value generation of data products is iterative. The value of data products is not solidified in a single instance; rather, it is formed through multiple rounds of refinement to create a "value-added data chain". For instance, user profiles evolve from basic labels (such as age and gender) to in-depth insights (such as consumption preferences and risk predictions) through multiple algorithmic iterations, with the rights object increasing in value dynamically. In addition, the application scenarios of data products are uncertain. Since the functions of data products are deeply tied to their application scenarios, they may lose economic value when detached from specific contexts. For example, health code data used during pandemic prevention and control may transform into data assets for urban governance research after the pandemic ends, marking a qualitative change in the rights object as the scenario shifts.

3.2 The Composite Nature of Rights Content

Traditional intellectual property rights exhibit linear characteristics in their content. For instance， copyright centers on the "right of reproduction,"extending to derivative rights such as distribution and communication through information networks, forming a bundle of rights focused on controlling utilization methods. Patent rights revolve around the "exclusive right to implement," encompassing specific entitlements like manufacturing, using, and selling, with clear and closed boundaries. Trademark rights prioritize the "exclusive right to use," supplemented by protections against consumer confusion and cross-class safeguards for well-known trademarks, all confined by designated product categories.

By contrast, the content of intellectual property rights in commercial data exhibits networked characteristics, specifically manifested as follows: (1) the multidimensional nature of affirmative rights. For example, based on the right to control, data processors have the authority to store and access data products (such as the classified and graded data management provisions in the *Data Security Law*); based on the right to develop, data processors may enhance data quality through algorithm optimization (such as the original labor involved in data cleaning and labeling); based on the right to benefit, relevant right holders are entitled to economic returns from data product transactions (such as licensing revenue-sharing mechanisms in data exchanges); (2) the defensive nature of negative rights. Data owners possess immunity to resist improper intervention by public authorities (such as commercial data exemption clauses in government data openness); they also hold the right to defense, allowing them to claim "fair use" in data scraping disputes (such as the "non-material substitution" defense in the *Dianping v. Baidu* case); (3) the interactive nature of interdependent rights. Interactivity is reflected in the coordination of conflicts between data intellectual property rights and individual information rights, such as the separation of data ownership and rights after anonymization processing; it also manifests in the convergence between data intellectual property rights and antitrust law, such as the determination of market dominance caused by data concentration [5].

3.3 Limitations on the Effect of Rights

Traditional intellectual property rights have strong exclusivity in their effect. For example, under copyright law, reproducing or distributing others' works without permission constitutes infringement, with exceptions strictly limited to statutory conditions—for instance, fair use must satisfy the "three-step test." Patent law grants right holders the authority to prohibit others from implementing identical or similar patented technologies, and independent development does not exempt one from infringement liability. Trademark law prohibits confusing use, with territorial protection strictly defining the boundaries of rights.

Different from traditional intellectual property rights, the effect of intellectual property rights in commercial data exhibits limited exclusivity. First, the scope of commercial data intellectual property rights is narrowed. Such protection excludes raw data and public data from rights assignment—for example, government-open data is not subject to data intellectual property rights. Additionally, the protection of derivative data products must meet an "innovative" threshold; the EU *Data Act*, for instance, requires data products to have "significant added value." Second, the exercise of commercial data intellectual property rights is restricted. During public crises (e.g., pandemic prevention and control) or for necessary technological innovation (e.g., AI training data), administrative authorities may compulsorily open commercial data usage rights; research institutions may use data for non-commercial purposes without authorization, as seen in the open-sharing principles of the *Regulations on Scientific Data Management*, which expand the concept of fair use in copyright law [6]. Furthermore, there are triggers for the invalidation of commercial data intellectual property rights. When data products lose economic value due to technological iteration, their rights automatically terminate—for example, outdated navigation data is no longer protected. When data subjects withdraw their consent, the rights to products derived from such data also become invalid.

In summary, the intellectual property rights in commercial data represent an extension of the intellectual property system in the digital era. In essence, they constitute a new type of information property rights, possessing both private and public attributes, and requiring institutional innovation to achieve the dual goals of incentivizing innovation and promoting circulation. Given that data continuously aggregates and evolves during its flow, the boundaries of derivative data products are uncertain, making the static object paradigm of traditional intellectual property rights difficult to adapt to the dynamic characteristics of data. It is necessary to respond to the needs of data mobility through flexible rights allocation, such as dynamic registration systems and scenario-based rights definition. Furthermore, the intellectual property rights in commercial data include not only affirmative rights (such as the right to use and license) but also negative defensive rights (such as immunity). The traditional "bundle of rights" model of intellectual property is insufficient to accommodate the complexity of data rights, necessitating the introduction of rights network theory to achieve dynamic balance of interests among multiple subjects. Finally, these rights must break through the "absolute control" logic of traditional intellectual property and adopt mechanisms of rights weakening and fair use to achieve Pareto optimality between data mobility and innovation incentives.

4 JUSTIFICATIONS FOR THE INTELLECTUAL PROPERTY PROTECTION OF COMMERCIAL DATA

4.1 Labor Theory of Property: Labor Input in Data Value

Locke's labor theory of property provides a moral-philosophical justificatory foundation for data rights assignment. The core proposition of this theory is that "labor removes natural objects from the common state and grants laborers exclusive rights." In the data domain, data processing labor exhibits significant heterogeneity, which can be divided into foundational labor and creative labor. Foundational labor includes low-skill repetitive tasks such as data collection, cleaning, and annotation. For example, data annotators classifying and labeling images, while time-consuming, lack originality; their labor value is mainly reflected in the "usability" of data resources. Creative labor, by contrast, involves high-intellectual activities such as algorithm design, model training, and knowledge discovery. For instance, using deep learning algorithms to extract disease diagnostic features from medical images—such labor creates new increments of knowledge, aligning with the core premise of Locke's theory that "labor creates value" [7].

Locke's labor theory of property provides a moral basis for data rights assignment. Data processors transform raw data into economically valuable derivative products through labor inputs such as data cleaning and algorithmic modeling. For example, e-commerce platforms generating consumer preference reports through user behavior data analysis—this process incorporates technical, financial, and intellectual labor, conforming to the justificatory logic of "labor creates property rights."

4.2 Utilitarian Theory: Incentivizing Innovation and Enhancing Efficiency

Bentham's utilitarian principle, which advocates for "the greatest happiness for the greatest number," manifests in data rights assignment as a need to balance innovation incentives and public welfare. First, granting data processors exclusive rights can motivate enterprises to increase investment in data R&D and promote data-driven technological innovation. However, excessive reinforcement of exclusive rights may also stifle subsequent innovation [8]. According to the patent thicket effect, if strong exclusivity is granted to data sets, downstream developers must obtain licenses at each layer, leading to a surge in transaction costs. For example, the "data silo" phenomenon in the medical data field forces AI drug development companies to spend years negotiating data usage rights. Additionally, over-strengthening exclusive rights carries the risk of inhibiting innovation—the EU *Database Directive*, which granted database creators a 15-year "sui generis" right, was criticized for "hindering data reuse," ultimately prompting the EU to shift toward an "open data strategy."

Thus, utilitarianism requires that rights allocation reserve institutional space for "maximizing overall social utility." Compulsory licensing mechanisms and data openness dividends reflect the balance between innovation incentives and public interests. When the marginal social benefit of data use exceeds the marginal cost of exclusive rights, the boundaries of rights protection should be transcended to grant the public certain open access rights [9].

4.3 Transaction Cost Theory: Internalization of Externalities

Coase's theorem posits that clear property rights definitions can reduce transaction costs. Applying this theorem requires reconstructing the efficiency boundaries of rights allocation in light of the unique characteristics of data as a factor of production. Data transaction costs include search costs, verification costs, and enforcement costs. Data transactions may also generate dual effects of positive and negative externalities: the former is exemplified by the open-source data platform Kaggle, which promotes global AI competition innovation through shared datasets and has incubated over 200,000 machine learning models; the latter, such as unanonymized geographic location data potentially used for crime prediction, triggering privacy disputes [1].

However, the prevalent information asymmetry and "Arrow's paradox" (disclosure leads to loss of value) in current data transactions cause market failures. Therefore, data intellectual property rights reduce uncertainties in data circulation by clarifying rights boundaries, internalizing economic externalities. For example, commercial data processors, assured of rights protection, are more willing to participate in data transactions, thereby promoting the optimal allocation of data resources.

5 APPROACHES TO THE INTELLECTUAL PROPERTY PROTECTION OF COMMERCIAL DATA

Despite the justificatory foundation for protecting data intellectual property rights, significant challenges remain. First is the ambiguity of rights boundaries. Pntellectual property protection for commercial data currently face dual risks of "overprotection" and "underprotection" : on the one hand, excessive emphasis on exclusivity may hinder data sharing and create "data silos"; on the other hand, unclear protection standards for derivative data in existing systems lead to heavy reliance on general clauses of the *Anti-Unfair Competition Law* in judicial practice, resulting in inconsistent adjudication standards[2]. Second is the complexity of interest conflicts. Data rights involve the interests of multiple stakeholders, posing significant challenges to balance: conflicts exist between private rights and public welfare, as data monopolies may threaten national security and public interests [10]; conflicts also arise between domestic laws and international rules, as divergent data governance models across countries impede cross-border data flows. Finally, there is the impact of technological iteration. Technologies such as blockchain and artificial intelligence (AI) challenge traditional intellectual property systems—for example, distributed ledger technology weakens the effectiveness of centralized registration systems, and existing regulations lack clear provisions on the originality of AI-generated data. In response to these legal challenges, this paper attempts to explore the intellectual property protection approaches for commercial data from the following aspects.

5.1 Building a Governance Model of "Limited Rights + Fair Use"

Given the significant challenges in protecting commercial data intellectual property rights, a governance model of "limited rights + fair use" can be constructed based on the rights attributes of commercial data intellectual property. The data industry chain involves multiple stakeholders: data contributors (individuals/enterprises), data processors (algorithm developers), and data users (consumers/third parties). Based on the theory of interest balancing, rights allocation should follow the "contribution principle," granting right holders limited rights [11].

On the one hand, in line with the value-added law of the data value chain, a hierarchical protection system of "raw data—basic data products—value-added data products" can be established: (1) Raw data (e.g., original temperature records collected by sensors), lacking creative labor input, should be regulated by laws such as the *Data Security Law* and *Personal Information Protection Law*; (2) Basic data products (e.g., cleaned structured datasets), although involving preliminary processing, have limited creativity and can be protected by property rights rules to safeguard the legitimate rights of data controllers; (3) Value-added data products (e.g., market demand prediction models formed through machine learning), which embody substantial intellectual creation and meet the dual standards of "original expression" and "manifest commercial value," should be included in the scope of intellectual property objects. This classification not only aligns with the "three-rights separation reform" direction proposed in the *Twenty Guidelines on Data* (i.e., separating data resource ownership, data processing and use rights, and data product operation rights) but also balances the dual needs of data circulation and rights protection.

Second, the limitation of rights can also be implemented through the protection term of data intellectual property rights. In general, the protection term for data intellectual property rights should be shorter than that of traditional intellectual property to promote data circulation. This paper proposes adopting a "statutory minimum protection period + dynamic adjustment" mechanism, namely: a basic protection period of 3–5 years for data intellectual property rights to ensure data processors recoup their innovation costs; for derivative data that is continuously updated and has significant social value, an extension may be applied for, with a maximum period not exceeding 10 years.

Additionally, fair use mechanisms or compulsory licensing systems should be applied to limit data intellectual property rights. For example, exceptions such as non-commercial use and scientific research can be established to ensure the

realization of data's public attributes. When public interests require it, administrative authorities may grant compulsory data usage licenses but must provide reasonable compensation.

5.2 Strengthening Technological Governance and Collaborative Regulation

First, the immutability feature of blockchain technology provides a new technical path for data rights attribution. By constructing a distributed ledger architecture, full-lifecycle behaviors such as data production, circulation, and processing are fixed in the form of hash values, forming a tripartite evidence chain that includes timestamps, operating subjects, and data fingerprints. This mechanism relies on asymmetric encryption algorithms (such as RSA, Elliptic Curve Cryptography) to digitally sign and authenticate data operation behaviors, ensuring that each data interaction can be traced and verified through the chain structure. In judicial practice, blockchain evidence storage that complies with the *Provisions of the Supreme People's Court on Several Issues Concerning the Trial of Cases by Internet Courts* has the effect of electronic evidence, effectively solving the problem of reversed burden of proof in data rights disputes. However, it should be noted that there is an inherent conflict between the storage efficiency of public blockchains and privacy protection, which requires realizing permission separation between regulatory nodes and business nodes through an alliance chain architecture [12].

Moreover, a collaborative regulatory system of "government-platform-third-party certification institutions" can be constructed. First, enable the docking of cross-platform data regulatory interfaces through interoperability standards. Second, use regulatory sandbox mechanisms to conduct risk stress tests on new data technologies. Third, introduce third-party audit institutions to carry out interpretability certification for algorithmic models.

5.3 Promoting International Rule Coordination

The international community currently faces a dilemma of "Regime Competition" in the protection of commercial data intellectual property rights: the WIPO *Treaty on the Protection of Databases* has been long dormant due to disagreements among member states; the TRIPS Agreement does not clarify the protection boundaries for data products; and while Chapter 18 of the CPTPP establishes the principle of free data flow, it lacks specific protection standards. To resolve this impasse, a "functionalism-oriented" governance framework should be constructed. In the first tier, relying on the WTO e-commerce negotiation mechanism to establish "technology-neutral" standards for determining the originality of data products. In the second tier, through the RCEP Agreement, establish non-discriminatory protection principles for derivative data products, and set up a dispute settlement panel for data intellectual property within the RCEP framework to build a specialized arbitration mechanism that includes a technical investigator system. In the third tier**,** exploring the convergence between domestic laws and international rules. Leverage the Guangdong-Hong Kong-Macao Greater Bay Area's cross-border data flow pilots and the RCEP Agreement to construct a unified protection framework for RCEP commercial data intellectual property rights, achieving collaborative governance of commercial data intellectual property within the RCEP region.

Resolving the dilemma of international rule coordination requires achieving two-way interaction between the "legalization of technical standards" and the "technicalization of legal rules." In the rights confirmation stage, under the RCEP framework, promoting the mutual recognition mechanism for data intellectual property registration procedures, incorporating elements such as blockchain evidence timestamps and data feature hash values into a unified certification system. In the rights exercise stage**,** translating the algorithmic interpretability requirements of GDPR Article 22 into enforceable technical standards to ensure that the use of data products complies with the transparency requirements of China's *Personal Information Protection Law* Article 24. In the rights protection stage, relying on the WIPO Lex database to establish a cross-border data intellectual property infringement characteristics library, using AI technology for real-time monitoring of cross-border infringement.

Additionally, the protection of commercial data intellectual property faces an inherent conflict between "legal territorialism" and "data borderlessness." Domestically, China needs to construct an "Embedded Compliance" framework. At the legislative level, through judicial interpretations of Article 127 of the *Civil Code* (data rights and interests clause), clarifying the "usufructuary rights of intellectual property" nature of derivative data products. At the judicial level**,** drawing inspiration from California's *Consumer Privacy Act* Section 1798.150 to establish a punitive compensation system for data intellectual property infringement (proposing a base amount of 1–3 times the actual losses of the right holder). At the enforcement level**,** relying on the National Data Administration to build a data intellectual property record-filing platform [2].

6 CONCLUSION

The legal attributes of commercial data and its intellectual property protection approaches need to break free from traditional paradigms and seek dynamic balance between incentivizing innovation and promoting circulation. By adopting a governance model of "limited rights + fair use," strengthening technological governance and collaborative regulation, and promoting international rule coordination, a data property rights system can be built that balances stability and adaptability. Future research should focus on the standardization of commercial data rights confirmation, judicial practices for rights relief, and the localized adaptation of international rules to provide a solid institutional foundation for the development of the digital economy.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Jia L P. Justification of Rights and Elaboration of Rules for Data Intellectual Property. Law and Social Development, 2024(4): 207–208, 213.
[2] Gao L. Dilemmas and Solutions in the Empowerment of Data Intellectual Property Rights. Jianghai Academic Journal, 2025(1): 185–186, 191.
[3] Liu X. Normative Approach and Institutional Framework for the Registration of Data Intellectual Property Rights. Journal of Huazhong University of Science and Technology (Social Sciences Edition), 2025(2): 92.
[4] Lv B B. An Intellectual Property Approach to the Establishment of Data Property Rights. Law and Business Review, 2025(1): 90.
[5] Rainer Kulms. Data Sharing and Data Protection. Romanian Review of Private Law, 2022(1): 151.
[6] Yanan Wang, Shanshan Cui. Deepening Digital Intellectual Property Rules and Digital International Patenting. Economics Letters, 2025, 251: 12.
[7] Zhang X B. On Data Property Rights as a New Type of Property Right. Social Sciences in China, 2023(4): 155–156.
[8] Feng X Q. Theoretical Explanation and Construction of Data Property Rights and Their Legal Regulation. Tribune of Political Science and Law, 2021(4): 83–84.
[9] Yang L H, Wang S T. Study on the Justification of Protecting Intellectual Property Rights in Commercial Data. Wuling Journal, 2024(1): 103–104.
[10] Wu H D. Empowerment of Data Property Rights: From Data Exclusive Rights to Data Use Rights. Law and Business Review, 2024(3): 10.
[11] Cui G B. Foundational Theory of Limited Exclusivity in Big Data. China Legal Science, 2019(5): 9.
[12] Eugenia Pedro, Joao Leitao, Helena Alves. Screening and Enhancing Intellectual Capital Consistency: A Scoping Review of Systematised Literature Reviews. Journal of Innovation & Knowledge, 2025, 10(2): 10.