

Article

From Fragmentation to Integration: The Structural Transformation and Maturation Mechanism of Data Factor Markets in China

Jiuxing Wu 

School of Geography and Tourism, Anhui Normal University, Wuhu 241002, China; wjx1211@ahnu.edu.cn

Abstract

Data has become a strategic production factor, but the institutional logic underlying data's tradability, priceability, and governability remains insufficiently theorized. In response, this study develops a coevolutionary framework that connects conventional factor market theory with digital political economy, platform theory, and comparative institutional analysis. This study adopts a conceptual–analytical research design, integrating three research methods: theory synthesis, comparative institutional analysis, and policy–process interpretation. Through theoretical synthesis, institutional comparison, and policy–process interpretation, it analyzes the conditions under which data circulation becomes feasible, lawful, and economically sustainable. In addition, by combining transaction data, exchange listings, property rights registrations, network indicators, and regional policy variations, it formulates testable propositions and an empirical agenda. The study finds that data factor markets do not emerge automatically with digitalization; their formation requires three mutually reinforcing conditions: technologically reducing search, verification, privacy protection, and contract enforcement costs; institutionally realizing a modular definition of rights and establishing compliance boundaries; and market demand from firms, public agencies, and research organizations generating use-case-specific value. Meanwhile, this study revises the three-stage model of market evolution as a contingent and testable pathway—from administrative pilot allocation, through hybrid state–market professionalization, to ecosystem-based cross-domain circulation. It also clarifies a closed-loop dynamic mechanism consisting of external shocks, internal strategic feedback, and adaptive governance, which jointly shapes market boundaries, pricing rules, and competition patterns.

Keywords: data factors; data factor market; formation mechanism; evolution mechanism



Received: 28 January 2026

Revised: 30 May 2026

Accepted: 23 June 2026

Published: 4 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The digital economy has altered the basic conditions under which production factors are accumulated, exchanged, and governed. Classical factor markets for land, labor, and capital are built around relative scarcity, exclusive possession, standardized transfer rules, and relatively mature price signals. Data differs from these factors because it is replicable, only partially rivalrous, often generated through multi-party interaction, and valuable mainly when linked with algorithms, computing infrastructure, and application scenarios. These properties make data an important source of productivity growth, but they also make the conversion of data into a marketable factor institutionally difficult (Goldfarb & Tucker, 2019; OECD, 2019; Acemoglu et al., 2022).

China provides a significant case for studying this transformation. Since data was formally identified as a factor of production in national policy, China has promoted data exchanges, public data authorization, rights registration, data merchants, data product listings, and the construction of data infrastructure. The 2022 policy package, often referred to as the Data 20 Articles, proposed a rights-division approach, and the National Data Administration was inaugurated in 2023 to coordinate Digital China, the digital economy, and data infrastructure development (CPC Central Committee & State Council of the People's Republic of China, 2022). These reforms have created a laboratory for observing how data moves from fragmented internal resources toward an organized factor market.

Literature has produced valuable insights into data ownership, data pricing, privacy-preserving computation, platform competition, and the macroeconomic effect of digital transformation (Koutroumpis et al., 2020; Azcoitia & Laoutaris, 2022; Driessen et al., 2022; Guo et al., 2025). However, three limitations remain. First, many studies treat technology, regulation, and demand as separate drivers rather than explaining how they reinforce or constrain one another. Second, the connection between conventional factor market theory and digital political economy remains underdeveloped: data markets are neither ordinary commodity markets nor purely administrative allocation systems. Third, stage models of market development are often presented as linear descriptions, while the conditions under which stages emerge, fail, or reverse are rarely specified in testable form.

This article addresses these gaps by developing a coevolutionary framework of data factor market formation and maturation. The term coevolutionary is used deliberately. It means that market rules, technological capabilities, participant strategies, and demand conditions adjust to one another over time. The data factor market is therefore not simply built by policy decree or released by technological innovation; it is produced through repeated interaction among institutional design, platform organization, and strategic behavior. This framing also prevents the three-stage model from being read as an inevitable reform sequence: each stage is treated as a hypothesis about observable coordination problems, institutional thresholds, and market responses.

To make the analytical promise of the article explicit, the study is organized around four research questions:

- RQ1: How do data factor markets differ from conventional factor markets, and what theoretical framework can explain this difference?
- RQ2: Through what technology–institution–demand mechanisms does data become tradable, priceable, and governable?
- RQ3: What stage transitions and feedback loops characterize the evolution of China's data factor market, and how can these claims be tested empirically?
- RQ4: How can the framework be translated into practical guidance for policy design, data exchanges, and sectoral applications?

The article makes four contributions. First, it connects conventional factor market theory with digital political economy and platform theory. Second, it integrates transaction cost theory, property rights theory, and network externalities into a dynamic mechanism rather than listing them as independent perspectives. Third, it reformulates the three-stage evolution of China's data factor market as a contingent, falsifiable model. Fourth, it provides hypotheses, measurable indicators, and empirical strategies that future researchers can use to test the model rather than treating it as an inevitable narrative.

2. Materials and Methods: Conceptual Research Design and Empirical Agenda

2.1. Research Logic and Source Base

The research design has four steps. First, it synthesizes three bodies of theory: transaction cost economics, property rights theory, and network externality/platform theory. Second, it compares data factor markets with classical factor markets to identify what is structurally new about data as a factor. Third, it interprets China's policy and market practice through process tracing, focusing on the movement from data trading pilots to data exchanges, rights registration, data merchants, public data authorization, and cross-domain data circulation. Fourth, it formulates propositions and empirical indicators that can be tested using market and policy data.

The evidence base is interpretive and documentary. It includes published studies on data markets and digital platforms, major data governance policies in China and the European Union, documented practices of data exchanges and data spaces, and empirical literature on data factor marketization, enterprise digital transformation, and platform competition. This combination supports analytical generalization: the article does not infer population-level causal effects, but it identifies mechanisms and scope conditions that can be evaluated in future empirical work. The article therefore positions its propositions as a research agenda rather than as already-tested causal estimates.

2.2. Validity Strategy and Limitations

To improve analytical validity, the study uses three checks. First, theoretical triangulation links each mechanism to more than one theoretical tradition. Second, comparative institutional analysis contrasts China's state-guided approach with the European access-and-data-space approach and the more platform-led, sectoral approach associated with the United States. Third, the empirical agenda specifies variables, identification strategies, and observable implications for each proposition. The main limitation is that the article remains conceptual; its propositions require future testing with transaction-level, firm-level, city-level, and platform-level data. To reduce the gap between conceptual theory and empirical research, Section 2.3 specifies hypotheses, indicators, and possible identification strategies that can be used to examine the framework.

2.3. Testable Propositions and Empirical Strategy

The conceptual framework yields four testable propositions. H1: stronger rights delineation, measured by property rights registration, standardized contracts, and compliance-review capacity, is associated with higher transaction frequency and more diversified data product listings. H2: transaction-cost-reducing technologies, such as trusted audit trails and privacy-preserving computation, increase cross-organization data circulation when combined with clear liability allocation. H3: network effects improve matching efficiency at early stages of market development but may reduce contestability when supplier or intermediary concentration becomes excessive. H4: regions or sectors with stronger demand-side application scenarios move faster from raw data transactions to data products, APIs, analytics services, and use-right contracts. These propositions can be examined using data-exchange transaction records, listed data products, rights registration records, regional policy rollouts, firm-level digital-transformation outcomes, and platform network indicators. Difference-in-differences designs, event studies, double machine learning, and network analysis are possible empirical strategies, but they should be applied only where data availability and identification assumptions are transparent.

3. Theoretical Foundations: Data as a New-Type Factor Market

3.1. From Classical Factor Markets to Data Factor Markets

Conventional factor markets are organized around scarcity, transferability, and enforceable exclusion. Land markets depend on spatially fixed property boundaries; labor markets depend on contracts and skills; capital markets depend on financial claims and risk pricing. In all three cases, property boundaries, transaction objects, and valuation conventions are relatively stabilized. Data changes these assumptions. Its value is relational, non-depletive, and often uncertain before use. It is generated by subjects, collected by platforms or organizations, processed by algorithms, and monetized through downstream applications. A single ownership model is therefore poorly suited to data (Table 1).

Table 1. Differences between classical factor markets and data factor markets.

Dimension	Classical Factor Markets	Data Factor Markets	Implication for Governance
Economic basis	Scarcity, exclusivity, and transfer of use or ownership	Replicability, partial non-rivalry, and value through recombination	Markets require access control and use-right governance rather than simple ownership transfer
Transaction object	Physical assets, labor services, or financial claims	Raw data, derived data products, models, APIs, analytics services, and usage rights	Pricing must be scenario-specific and risk-adjusted
Property boundary	Relatively stable and legally codified	Multi-party, layered, and context-dependent	Bundle-of-rights design is necessary
Value formation	Marginal productivity and scarcity-based allocation	Network effects, complementary assets, and algorithmic learning	Platform governance and interoperability shape value
Main risk	Search costs, contracting costs, opportunism	Privacy leakage, quality uncertainty, monopoly, compliance risk, cross-border restrictions	Markets require technical, legal, and organizational infrastructures

3.2. Digital Political Economy and Platform Theory

Digital political economy adds an important layer to factor-market theory. Data is not merely an input purchased in a neutral marketplace; it is often generated inside platform ecosystems where access, interoperability, visibility, and monetization are governed by platform rules. Platform theory shows that network effects can improve matching and innovation, but can also produce lock-in, market concentration, and asymmetric bargaining power (Rochet & Tirole, 2006; Parker et al., 2016; Crémer et al., 2019). This is why data factor markets are simultaneously allocation systems and governance systems.

The recent research on digital transformation and platform-enabled entrepreneurship strengthens this point. Digital transformation can improve enterprise productivity by reorganizing capabilities and entrepreneurship (Guo et al., 2025). AI adoption in online travel

platforms demonstrates how data, algorithms, and platform market structure co-determine pricing, competition, and value capture (Hernández-Tamurejo et al., 2025). Studies of social media, crowdsourcing, and market-oriented entrepreneurship show that data-driven participation can support organizational learning, but only when knowledge flows are structured and absorbed by firms (Juárez-Varón & Monreal, 2025). These contributions suggest that a data factor market is not just a venue for exchanging datasets; it is a platform-mediated institutional arrangement for organizing data-based learning and value creation.

This connection between factor-market theory and digital political economy is especially important because data markets combine allocation, governance, and platform control. Classical factor theory explains why stable rights, transaction objects, and price signals matter. Digital political economy explains why those elements are shaped by platform power, datafication, algorithmic intermediation, and state–market relations. Comparative institutional analysis then clarifies why China’s state-guided data-exchange model, the European dataspace and access-rights model, and more platform-led sectoral models generate different combinations of circulation, compliance, and competition risk.

3.3. Multi-Dimensional Theoretical Core and Dynamic Mechanism of Data Factor Markets

Understanding data as a production factor requires an integrated theoretical framework that links its intrinsic economic attributes with the institutional mechanisms through which data become tradable, scalable, and governable. Unlike traditional factors such as land, labor, and capital, data possess quasi-public good characteristics, relative non-rivalry and institutionally constructed excludability, and strong network effects. These attributes challenge classical assumptions of market exchange and explain why data factor markets require specialized governance arrangements, rights delineation mechanisms, and platform-based transaction infrastructures.

First, data exhibit quasi-public good characteristics. Although data production requires substantial upfront investment in collection, storage, processing, and security infrastructure, the marginal cost of replication and reuse is extremely low. Thus, data use is largely non-rivalrous, as one party’s use does not necessarily diminish another’s ability to use the same dataset. At the same time, modern access-control technologies, encryption protocols, contractual restrictions, and regulatory mechanisms enable partial excludability, making data neither purely public nor purely private, but rather a hybrid quasi-public good (Goldfarb & Tucker, 2019; OECD, 2019). This hybrid nature is especially salient in high-value datasets such as healthcare records, transportation data, environmental monitoring data, and public administrative data, whose circulation affects public service efficiency, innovation capacity, and social governance. Consequently, data factor markets must balance efficiency-oriented data monetization with public-interest constraints, including privacy protection, ethical use, and national data security (European Commission, 2020; Acemoglu et al., 2022).

Second, the non-rivalry and excludability of data are relative rather than absolute. While data can technically be reused by multiple actors, congestion may arise when access demand exceeds the capacity of storage systems, computing resources, or data-interface services, thereby creating practical rivalry in availability and processing performance (Belleflamme & Peitz, 2019). Moreover, excessive duplication and uncontrolled circulation may undermine data integrity, reliability, and commercial scarcity. Conversely, data excludability has become increasingly institutionalized through encryption, secure multi-party computation, federated learning, trusted execution environments, contractual licensing, and regulatory regimes such as the EU General Data Protection Regulation, China’s Data Security Law, and emerging data licensing frameworks (European Union, 2016; Gao & Shao, 2021). Therefore, data’s marketability depends less on natural scarcity than on governance

infrastructure capable of defining access, use, processing, transfer, benefit-sharing, and revocation boundaries.

Third, data factor markets are shaped by strong network effects. The value of data increases with the number of participants, the scale of data accumulation, and the diversity of interconnections. Larger and more diverse datasets improve statistical inference, algorithmic accuracy, and predictive performance, thereby raising the marginal value of additional data contributors and users (Brynjolfsson et al., 2018). Isolated datasets often have limited analytical value, whereas multi-source data aggregation generates complementary insights and innovation spillovers. These dynamics create endogenous incentives for data pooling, platform formation, interface standardization, and ecosystem-based organization. As participation expands, positive feedback loops may improve matching efficiency and accelerate value creation, but they may also reinforce scale advantages, platform concentration, and ecosystem lock-in (Rochet & Tirole, 2006; Parker et al., 2016).

On this basis, transaction cost theory, property rights theory, and network externality theory jointly explain the formation and maturation of data factor markets (Figure 1). Transaction cost theory clarifies why organized data markets emerge in response to the inefficiencies of informal, bilateral, or relational data exchange. Before formal market organization, data transactions are constrained by high search costs, unverifiable data quality, privacy and compliance risks, contractual incompleteness, and enforcement uncertainty (Williamson, 2010; Bajari & Tadelis, 2011). Data exchanges, data merchants, registries, compliance audits, standardized data descriptions, pricing mechanisms, trusted intermediaries, certification systems, and dispute resolution protocols reduce these frictions by improving information transparency, simplifying contracting, and enhancing enforcement credibility (OECD, 2019; J. Chen et al., 2021). Nevertheless, reducing transaction costs alone cannot guarantee effective data circulation, because data entitlements remain layered, contested, and context-dependent.

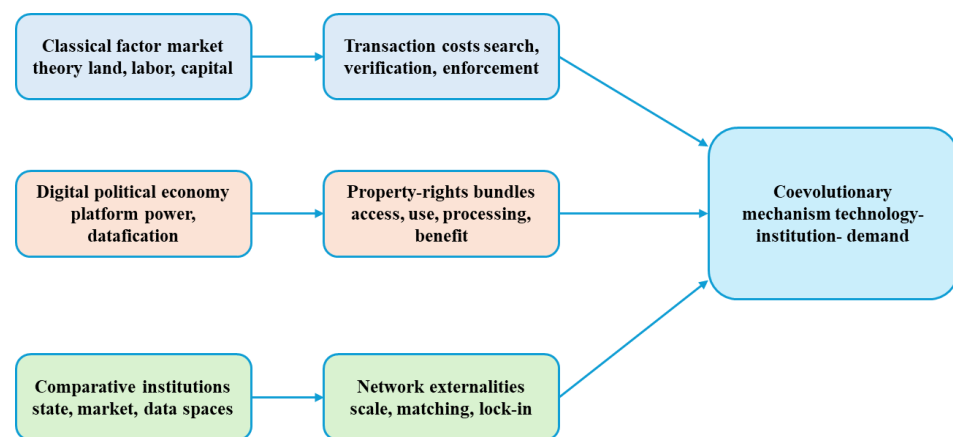


Figure 1. Integrated theoretical framework for data factor market formation and maturation.

Property rights theory therefore explains how data become marketable without requiring absolute ownership transfer. Data production usually involves multiple stakeholders, including data subjects, collectors, processors, analysts, platform operators, and application developers. At the same time, data embody personal rights, commercial interests, and public-interest concerns, making singular ownership assignment both impractical and normatively insufficient (Posner & Weyl, 2017; Acemoglu et al., 2022). Emerging governance frameworks thus adopt a bundle-of-rights approach, decomposing data entitlements into holding, access, processing, use, operation, revenue-sharing, and revocation rights. These rights are implemented through contractual licenses, technical access controls, compliance systems, and regulatory rules rather than through full ownership transfer (Hoofnagle et al.,

2019). This modularization allows different governance arrangements for public data, enterprise data, personal information, and sensitive data. Although China's rights-division approach and Europe's access-rights approach differ institutionally, both move beyond a simple ownership model and emphasize controlled, conditional, and accountable data use.

Network externality theory further explains why data factor markets tend toward scale expansion and platformization (Figure 2). In two-sided data markets, an expanding supply side enriches data variety and improves matching efficiency for demanders, while a growing demand side expands application scenarios, monetization opportunities, and incentives for further data provision (Parker et al., 2016; Rochet & Tirole, 2006). These bidirectional feedback mechanisms promote higher-frequency circulation and ecosystem-based innovation. However, they also generate risks of winner-takes-most market structures, bargaining asymmetry, switching costs, and platform lock-in. Therefore, the governance objective is not merely to promote data circulation, but to ensure that circulation remains contestable, interoperable, privacy-preserving, and aligned with public-interest constraints. Regulatory frameworks must balance innovation incentives with competition safeguards, data-sharing obligations, and mechanisms preventing excessive concentration (OECD, 2021; Crémer et al., 2019).

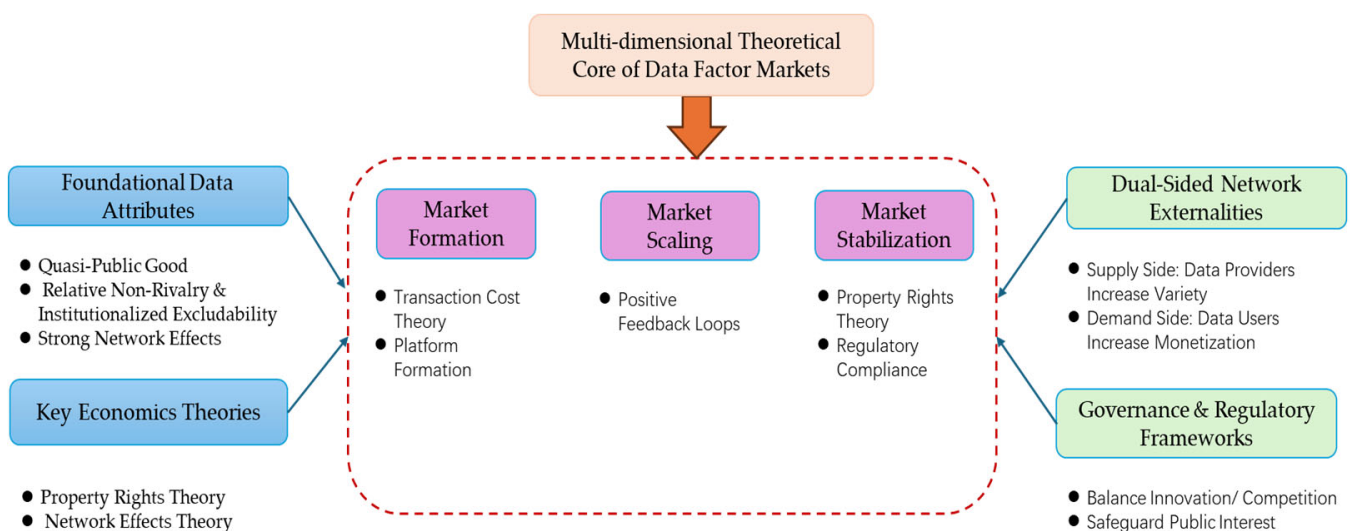


Figure 2. Multi-dimensional theoretical core of data factor markets.

The interaction among transaction costs, property rights, and network externalities is dynamic rather than sequential. Lower transaction costs increase market participation, but broader participation intensifies the need for modular rights allocation and compliance governance. Clearer rights reduce legal uncertainty and support tradability, but they also shape who can access data, under what conditions, and who captures network-generated value. Network externalities improve matching, learning, and innovation, yet they may simultaneously raise switching costs, strengthen incumbents, and increase bargaining asymmetries, thereby requiring renewed institutional adjustment. Accordingly, the formation of data factor markets should be understood as a recursive process in which technological capacity, rights structure, transaction governance, and network organization continuously reshape one another.

3.4. Three Perspectives on Data Factor Market Research

Technological innovation is widely viewed as a key driver of the emergence and expansion of data factor markets, that is, markets in which data is allocated and traded as an input for production and innovation. A prominent strand of research examines how

technologies such as blockchain, privacy-preserving computation, and big data analytics are applied to data transactions (Xu et al., 2023; M. Zhang et al., 2023; Yu et al., 2024). These tools are argued to mitigate information asymmetry and strengthen security and privacy safeguards, which may lower verification, contracting, and enforcement costs and facilitate lawful data exchange. For example, blockchain's decentralization and tamper resistance can improve auditability, support traceability of data flows, and enhance the enforceability of data-related rights (Yu et al., 2024; Li et al., 2023; Deepa et al., 2022). "Privacy computing" (an umbrella term for methods that enable analysis without exposing raw data) can allow cross-organization data use while reducing disclosure risks (Xu et al., 2023; Yu et al., 2024), thereby easing constraints associated with data silos.

Institutional regulation highlights the enabling and protective role of governance arrangements in these markets. Research in this area addresses property rights allocation, security and privacy rules, transaction standards, and oversight mechanisms (Duch-Brown et al., 2017). Many studies contend that well-specified rights, coherent legal frameworks, and transparent trading rules are prerequisites for market stability and participation. One proposal is the "separation of three rights" approach, which distinguishes ownership, usage rights, and the right to obtain economic benefits (usufruct), with the aim of balancing stakeholder interests and improving the efficiency of data allocation. Other work analyzes how national data security regimes shape market behavior, arguing that appropriately designed regulation can reduce compliance uncertainty, contain systemic risk, and strengthen market confidence (Ke & Sudhir, 2022; Margoni et al., 2023; Yu et al., 2024).

A third perspective focuses on market structure, including participant roles, competitive dynamics, and pricing mechanisms. Studies examine interactions among data suppliers, intermediaries (e.g., brokers, platforms, and service providers), and data users, with particular attention to entry barriers and monopoly risks. Some scholars argue that data markets exhibit a "Matthew effect" (cumulative advantage) (Abrardi et al., 2024a, 2024b), whereby large technology firms leverage scale and existing data advantages to consolidate market power and weaken competition. Pricing research similarly notes that data value is context-dependent and uncertain, limiting the applicability of traditional factor-pricing approaches; accordingly, researchers propose dynamic pricing methods grounded in use-case-specific value assessment and risk-adjusted valuation (Xu et al., 2023; Y. Chen et al., 2024).

4. Formation Mechanism: Technology—Institution—Demand Synergy

The formation of a data factor market depends on the joint operation of three mechanisms. Technology expands the feasible set of data circulation. Institutions define lawful and credible boundaries for exchange. Demand turns technical and legal possibilities into recurring market activity. These mechanisms are mutually reinforcing: technology without institutions produces risk; institutions without demand produce formal rules without liquidity; and demand without technology and institutions produces fragmented and unsafe exchange.

4.1. Formation Drivers: A Multi-Dimensional Driving Model

Overall, the formation of the data factor market is not the result of a single factor, but rather a synergistic product of technological, institutional, and economic multi-dimensional factors (Figure 3).

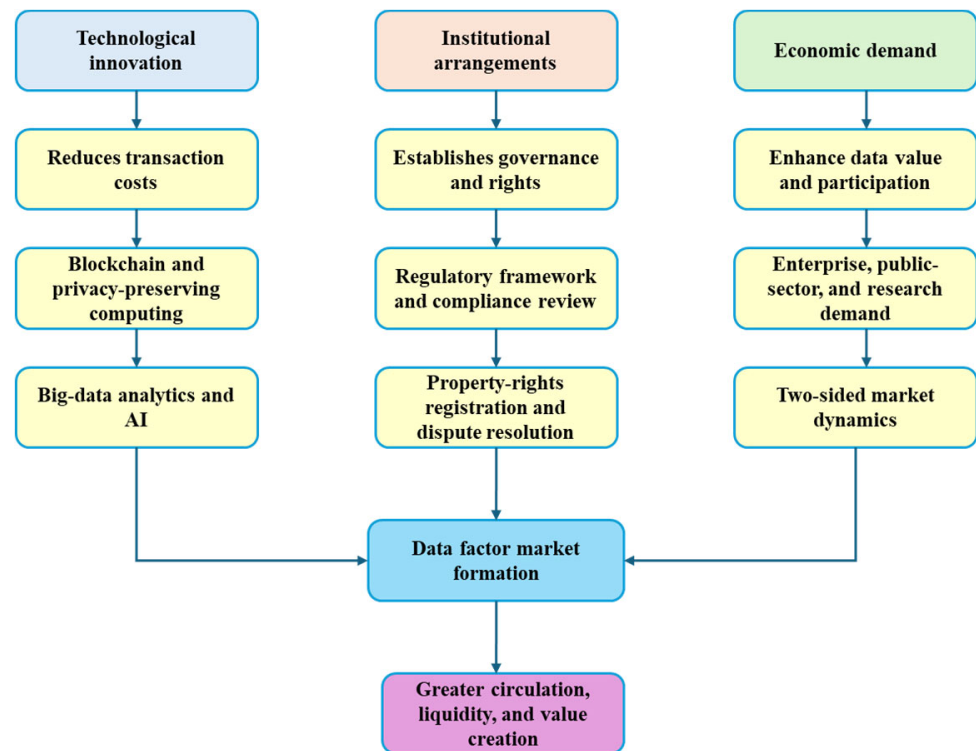


Figure 3. Technology–institution–demand formation mechanism of the data factor market.

4.1.1.1. Technological Innovation Reduces Transaction Costs

Technological innovation reduces several transaction costs specific to data. Blockchain and tamper-resistant audit logs can support traceability, authorization records, and settlement. Privacy-preserving computation, including secure multi-party computation, federated learning, trusted execution environments, and differential privacy, enables value extraction under controlled conditions. Data cleaning, labeling, metadata management, and quality certification convert raw data into usable data products. AI systems increase the value of multi-source data by improving prediction, classification, and decision support. In terms of mechanisms, technology shifts the feasibility frontier: it makes transactions possible that would otherwise be blocked by privacy, security, or verification constraints. The innovation and application of digital technologies such as blockchain, privacy computing, big data analytics, and artificial intelligence are the core technological drivers for the formation of the data factor market. These technologies address key pain points in data transactions, effectively reducing transaction costs and providing technological support for the market's emergence (Lu et al., 2020).

Blockchain technology provides a decentralized trust mechanism for data transactions (H. Wang, 2025). In traditional data transactions, trading parties rely on third-party institutions to establish trust, increasing transaction costs and enforcement risks. Blockchain's decentralized, immutable, and traceable characteristics enable full transaction traceability and immutability, reducing information asymmetry and lowering trust and enforcement costs. For instance, blockchain can be used for data property rights registration and certification, clarifying ownership and providing a legal basis for data transactions (H. Chen, 2023). It can also facilitate data transaction clearing and settlement, improving efficiency and reducing costs.

Privacy-preserving computing technologies address data security and privacy protection constraints in data transactions. Extracting value from data factors often requires the aggregation and sharing of multi-party data. However, data sharing can lead to data leaks and privacy infringements, which are critical bottlenecks constraining data transac-

tions. Privacy computing technologies (e.g., federated learning, differential privacy, secure multi-party computation) enable joint analysis and value extraction from data without exposing raw data, breaking down data silos, and promoting data factor circulation and sharing. For example, in medical data transactions, federated learning allows various medical institutions to jointly train diagnostic models without disclosing original patient data, maximizing the value of medical data.

Big data analytics and artificial intelligence technologies enhance data value extraction capabilities. The value of data factors is often latent and requires professional analysis and extraction to be revealed. Big data analytics technologies can clean, organize, and analyze massive datasets to uncover hidden patterns and value. Artificial intelligence technologies enable automated data analysis and prediction, improving the efficiency and accuracy of data extraction (D. Wang & Yang, 2025). The application of these technologies increases the value of data factors, boosts market participants' enthusiasm for data transactions, and drives the formation and development of the data factor market.

4.1.2. Institutional Arrangements Provide Governance Safeguards

Institutional supply transforms possible exchange into legitimate exchange. The most important institutional functions are rights delineation, compliance review, security classification, standard setting, dispute resolution, and competition oversight. In China, the Data 20 Articles emphasize a rights-division logic that distinguishes data resource holding rights, data processing and use rights, and data product operation rights. In the European Union, the Data Governance Act and Data Act use a different route by strengthening trusted data-sharing mechanisms, access rights, and rules for connected product and service data. These approaches differ, but both show that data factor markets are rule-intensive markets.

Institutions are the foundation for regulating market behavior and ensuring orderly market operation. A comprehensive policy supply and institutional system are crucial institutional drivers for the formation of the data factor market. In recent years, China and major countries worldwide have successively introduced a series of policy documents focusing on key issues such as data property rights, data security, and data transaction rules, constructing an institutional framework for data factor market development and providing institutional guarantees for its formation.

Data property rights systems provide a legal and operational basis for data transactions. Clear data property rights are a prerequisite for data transactions (L. Chen, 2023; J. Huang, 2024; L. Huang et al., 2021). If property rights are ill-defined, it leads to unclear responsibilities and frequent disputes, hindering transactions. Opinions of the Central Committee of the Communist Party of China and the State Council on Establishing Fundamental Data Institutions to Better Leverage the Role of Data as a Factor of Production (Abbreviation as "Data 20 Articles") proposed establishing a "separation of three rights" system for data property, clarifying the ownership and boundaries of data resource holding rights, data processing and usage rights, and data product operating rights, providing fundamental guidance for data property rights delineation. Concurrently, various regions are actively exploring data property rights registration and certification mechanisms, such as the Shanghai Data Exchange establishing a data property rights registration platform to provide proof of ownership for data transactions, effectively reducing legal risks.

Data security and privacy regulations delineate safety boundaries for data transactions. Data security and privacy protection are the bottom line for the healthy development of the data factor market. Without adequate data security, market participants' confidence will be severely undermined. China has successively introduced laws and regulations such as the "Cybersecurity Law," "Data Security Law," and "Personal Information Protection Law," establishing a legal framework for data security and privacy protection. The "Data

20 Articles” further clarify responsibilities and obligations for data security protection, proposing a data classification and grading protection system to apply differentiated protection measures to different types and levels of data. The promulgation of these laws and regulations sets safety boundaries for data transactions, reduces security risks, and enhances market participants’ confidence.

Data transaction rules provide operational norms for data transactions. The unique nature of data transactions necessitates specialized transaction rules to regulate behavior. Data exchanges across China have actively explored and formulated data transaction rules. For example, the Guiyang Big Data Exchange has formulated data transaction management measures and operational norms, specifying participants, processes, pricing mechanisms, and risk control. The Shanghai Data Exchange innovatively proposed a “data merchant system,” constructing a full-chain transaction ecosystem covering data suppliers, intermediary service providers, and demand-side parties, and formulating rules for merchant access, evaluation, and exit, enhancing the standardization of data transactions (D. Wang & Yang, 2025; Anand et al., 2025). The refinement of these transaction rules provides operational guidelines, reduces transaction costs, and promotes the efficient circulation of data factors.

4.1.3. Economic Demand Converts Data into Use-Case Value

Demand gives data marketization its economic direction. Enterprises seek data to improve product design, risk control, logistics, marketing, and intelligent production. Public agencies seek data to improve urban governance, public service delivery, and emergency response. Research institutions seek access to datasets for scientific discovery and model training. Because data value is context-specific, demand is not a simple volume variable. Data becomes valuable when it is linked to a specific decision, model, product, or service. This explains the shift from raw data transactions toward data products, APIs, analytics reports, model services, and use-right contracts.

The demand for data factor value realization and the need for improved overall societal resource allocation efficiency are the core economic drivers for the formation of the data factor market. With the development of the digital economy, the value of data factors is increasingly prominent, and market entities’ demand for data factors is continuously growing. However, traditional data allocation methods struggle to meet market demand, necessitating the construction of a data factor market to achieve market-based allocation, unlock data value, and enhance resource allocation efficiency (R. Chen, 2023; Duan et al., 2025).

Firstly, enterprise demand for data factors is the micro-foundation for the data factor market’s formation. In the digital economy era, data factors have become a critical resource for enterprises to enhance their core competitiveness. By acquiring and analyzing data, enterprises can precisely understand market demands, optimize product design, improve production efficiency, and reduce operational costs. For example, e-commerce companies can achieve precise marketing by analyzing user consumption data, improving marketing efficiency. Manufacturing companies can optimize production processes and reduce costs by analyzing production data. As enterprise demand for data factors continuously grows, traditional decentralized data acquisition methods can no longer satisfy enterprises’ large-scale, diversified data needs. The urgent need to acquire stable, high-quality data resources through a data factor market provides demand-side support for its formation.

Secondly, the demand for improved overall societal resource allocation efficiency is the macro-driver for the data factor market’s formation. Traditional resource allocation primarily relies on market mechanisms and government regulation, but these have certain limitations in resource allocation efficiency (Yuan & Du, 2025). Data factors, with their unique attributes of non-rivalry and network effects, can break through the spatial and

temporal constraints of traditional factors, enabling cross-regional and cross-industry optimization of resources. By constructing a data factor market, data factors can freely circulate and be optimally allocated across society, promoting the aggregation of production factors in high-efficiency areas and enhancing overall societal resource allocation efficiency. For example, the opening and sharing of government data can drive the optimization of government service processes, improving public service efficiency. The circulation and sharing of industry data can promote collaboration within industrial chains, driving industrial upgrading.

4.2. Market Evolution Stages: From Fragmentation to Systematization

The market evolution stages model identifies thresholds and transition conditions rather than a deterministic timeline. The formation of the data factor market is a gradual development process, following an evolutionary path from fragmentation to systematization, which can be divided into three stages: primary, intermediate, and advanced (Figure 4). The three-stage model is useful only if it is treated as a contingent pathway rather than as an inevitable sequence. A region or sector may remain in the pilot stage if rights are unclear, demand is weak, or privacy risks are unresolved. A platform may move rapidly toward ecosystem integration if it combines strong technical infrastructure, standardized contracts, large demand, and trusted intermediaries.

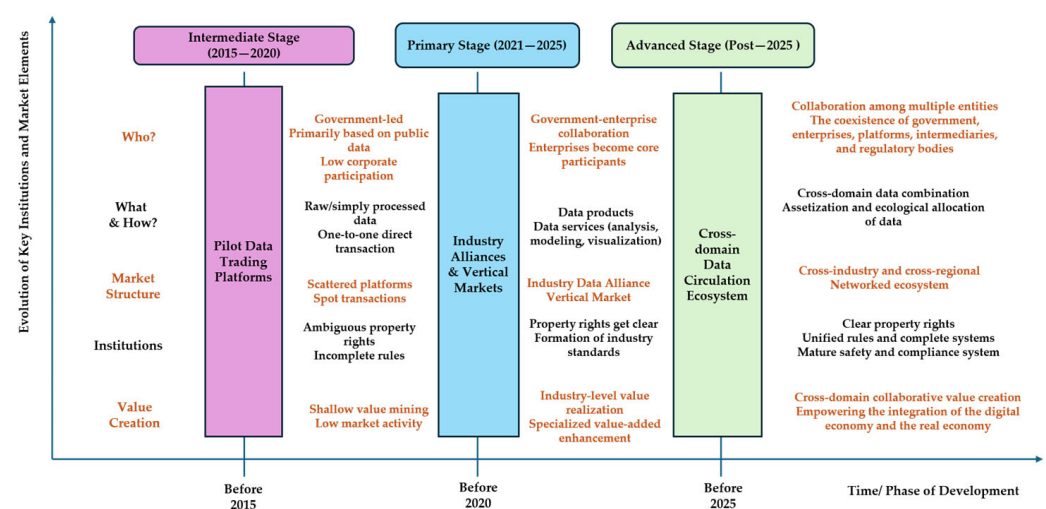


Figure 4. Contingent evolutionary pathway of the data factor market.

International experience suggests that the three-stage model should be treated as a comparative and contingent framework. The European Union emphasizes trusted data intermediaries, data spaces, and access-right governance; platform-led systems emphasize private ecosystem control, APIs, and contractual access; China emphasizes public policy coordination, data exchanges, rights registration, and public data operations. These pathways differ, but all face the same threshold problems: defining lawful access, lowering transaction risks, building demand liquidity, and preventing excessive concentration.

4.2.1. Primary Stage

The primary stage of data factor market formation is characterized by pilot data trading platforms, roughly from 2015–2020. In this stage, the production factor attribute of data gradually gained recognition, and policies began to encourage the construction of data trading platforms. Various regions launched pilot data trading platforms to explore transaction models and pathways. The main characteristics of this stage are: Firstly, single transaction entities, primarily government-led public data transactions, with a low propor-

tion of enterprise data transactions. Most data trading platforms were government-led, focusing on the opening and trading of government data, such as traffic data, meteorological data, and government service data (Duan et al., 2025; Ye et al., 2022; D. Wang & Yang, 2025). Enterprises, lacking clear property rights definitions and transaction rules, showed low enthusiasm for participation. Secondly, simple transaction models, mainly direct transactions of data resources, with a low proportion of data product and service transactions. Trading parties primarily adopted “one-to-one” direct transaction models, dealing mainly with raw or simply processed data, with a low degree of data value extraction (Duan et al., 2025; Liang et al., 2018; Ye et al., 2022). Thirdly, small transaction scale and low market activity. Due to constraints such as unclear data property rights, imperfect transaction rules, and high data security risks, the scale of data transactions was small, and market activity was low (Ye et al., 2022).

4.2.2. Intermediate Stage

With the continuous advancement of data trading platform pilots, the data factor market gradually entered an intermediate stage, characterized by the formation of industry data alliances and vertical markets, roughly from 2021–2025. In this stage, policies further refined the institutional framework for the data factor market, data property rights delineation became clearer, and enterprises’ enthusiasm for participating in data transactions continuously increased. Concurrently, due to the varying characteristics and demands of data across different industries, industry data alliances and vertical markets gradually formed, enhancing the professionalization of data transactions. The main characteristics of this stage are: Firstly, diversified transaction entities, with enterprises becoming core participants in data transactions. As data property rights delineation continuously improved and transaction rules became clearer, enterprises’ enthusiasm for participating in data transactions increased (D. Wang & Yang, 2025; Duan et al., 2025). Transaction entities gradually shifted from government-led to collaborative participation by government and enterprises, with the proportion of enterprise data transactions continuously rising. Secondly, diversified transaction models, with a continuously increasing proportion of data product and service transactions. Trading parties were no longer limited to direct transactions of raw data but increasingly engaged in transactions of data products and services, such as data analysis reports, data modeling services, and data visualization services, leading to a continuously higher degree of data extraction (Liang et al., 2018; Ye et al., 2022). Thirdly, the rise of industry data alliances and the gradual formation of vertical markets. Enterprises in different industries formed industry data alliances to achieve data resource sharing and collaboration, such as financial data alliances, medical data alliances, and industrial data alliances. These alliances, by formulating industry data standards and standardizing data transaction processes, promoted the efficient circulation of data factors within industries, forming specialized vertical data markets (Jarke et al., 2019; Magas & Kiritsis, 2021).

4.2.3. Advanced Stage

The advanced stage of data factor market formation is characterized by the construction of cross-domain data circulation ecosystems, roughly post-2025. In this stage, the institutional framework of the data factor market is largely complete, data property rights are clearly defined, transaction rules are standardized, and data security is effectively guaranteed. Concurrently, with the development of new technologies such as AI large models and quantum computing, cross-domain circulation and integrated application of data factors become possible, and cross-domain data circulation ecosystems gradually take shape. The main characteristics of this stage are: Firstly, data factors achieve free

cross-domain circulation. Data barriers between industries, regions, and departments are broken, enabling data factors to circulate freely and be optimally allocated across society. For example, data from different domains such as government, finance, medical, and transportation can be integrated and applied, providing data support for the development of new business models and industries like smart cities, smart healthcare, and smart transportation (Jarke et al., 2019; Möller et al., 2024). Secondly, a complete data transaction ecosystem and sound intermediary service system. A complete data transaction ecosystem is formed, encompassing multiple participants such as data suppliers, intermediary service providers, demand-side parties, and regulatory agencies. Intermediary service providers offer professional services such as data cleaning, data labeling, data evaluation, legal consulting, and insurance, providing full-chain support for data transactions (Liang et al., 2018; Azcoitia & Laoutaris, 2022). Thirdly, data value is maximized. Through the integrated application of cross-domain data, the deep value of data factors is extracted, promoting the deep integration of the digital economy with the real economy, and injecting sustained momentum into high-quality economic development (Trakadas et al., 2022; Yang, 2025).

4.3. Key Supporting Elements

The formation and development of the data factor market require the synergistic efforts of three critical supporting elements: technological infrastructure, institutional safeguard system, and cultivation of market entities. None of these can be omitted.

4.3.1. Technological Infrastructure

Technological infrastructure is the hardware foundation for the formation and development of the data factor market. It primarily includes data processing capabilities such as data labeling, cleaning, and storage, as well as technical support capabilities like blockchain and privacy computing (J. Zhang et al., 2024; Azcoitia & Laoutaris, 2022; Eichler et al., 2022). Extracting value from data factors requires high-quality data resources. However, raw data often suffers from incompleteness, inaccuracy, and non-standardization, necessitating processing methods like data labeling and cleaning to improve data quality. Data storage capability is fundamental to ensuring data security and accessibility (Deepa et al., 2022; Jarke et al., 2019). As data volume continuously expands, higher demands are placed on storage capacity, speed, and security.

Firstly, data labeling is a crucial step in enhancing data quality. Data labeling involves annotating and tagging raw data so that it can be recognized and understood by machine learning models. High-quality data labeling can improve the training effectiveness of machine learning models and unlock the potential value of data. Currently, China's data labeling industry has matured, with specialized labeling teams and service platforms providing customized data labeling services for various industries.

Secondly, data cleaning involves de-duplication, imputation, and error correction of raw data to improve its accuracy and completeness. The continuous development of data cleaning technologies effectively enhances data quality, providing reliable data resources for data transactions.

Thirdly, the innovation and application of data storage technologies are fundamental to ensuring data security and accessibility. As data volumes continuously expand, traditional storage technologies are often insufficient. Cloud storage and distributed storage technologies have emerged to address this. Cloud storage offers advantages such as large capacity, scalability, and low cost, providing efficient storage services for the data factor market. Distributed storage boasts high reliability and availability, ensuring data security and accessibility. Concurrently, the application of security technologies like data encryption and access control further enhances data storage security, reducing the risk of data leakage.

4.3.2. Institutional Safeguard System

The institutional safeguard system is the bedrock for the healthy development of the data factor market, primarily comprising mechanisms for property rights registration, compliance review, and dispute resolution (J. Zhang et al., 2024; Fast et al., 2023; Martens, 2024). A property rights registration mechanism clarifies data ownership, providing a legal basis for data transactions. A compliance review mechanism ensures that data transactions adhere to laws, regulations, and policy requirements, reducing transaction risks. A dispute resolution mechanism promptly addresses disputes arising during data transactions, safeguarding the legitimate rights and interests of both trading parties (Schweihoff et al., 2024; Creemers, 2022; Fernandez et al., 2020).

Firstly, a property rights registration mechanism is a prerequisite for data transactions. Establishing a data property rights registration platform to register and certify data property rights can clarify the ownership of data resource holding rights, data processing and usage rights, and data product operating rights, providing proof of ownership for data transactions. For example, the Shanghai Data Exchange has established such a platform, effectively reducing legal risks associated with data transactions. Furthermore, property rights registration mechanisms enable traceability and management of data property rights, enhancing the efficiency of data property protection.

Secondly, a compliance review mechanism is key to ensuring the security of data transactions. Establishing specialized compliance review agencies to comprehensively examine transaction entities' qualifications, data sources, data uses, and data security ensures that data transactions comply with relevant laws and regulations, such as the "Cybersecurity Law," "Data Security Law," and "Personal Information Protection Law." For instance, data exchanges across various regions have established compliance review teams to rigorously review data transaction projects, prohibiting illegal data transactions and ensuring the legality and security of data transactions.

Thirdly, a dispute resolution mechanism is vital for protecting the legitimate rights and interests of trading parties. Data transactions may involve property rights disputes, contract disputes, and data leakage disputes. Therefore, establishing an efficient dispute resolution mechanism to promptly resolve these issues is crucial. Dispute resolution mechanisms can employ various methods, including negotiation, mediation, arbitration, and litigation. Among these, arbitration offers advantages of efficiency, convenience, and specialization, making it suitable for resolving data transaction disputes. Establishing data transaction arbitration institutions and cultivating specialized arbitration talent can provide efficient dispute resolution services for data transactions, ensuring the orderly operation of the data factor market.

4.3.3. Cultivation of Market Entities

Market entities are the core participants in the data factor market. The differentiation and collaborative cooperation of data suppliers, intermediary service providers, and demand-side parties are crucial for the market's formation and development (Schweihoff et al., 2024). Data suppliers provide data resources, representing the supply side of the data factor market. Intermediary service providers offer specialized services, acting as a bridge connecting data suppliers and demand-side parties (Jarke et al., 2019). Demand-side parties acquire and utilize data resources, representing the demand side of the data factor market.

Firstly, the differentiation and cultivation of data suppliers are central to the supply side of the data factor market. Data suppliers can be categorized into public data suppliers and enterprise data suppliers. Public data suppliers include government departments and public institutions, responsible for providing government data and public service data. Enterprise data suppliers include internet companies, financial institutions, and

manufacturing enterprises, providing enterprise operational data and user data. Through policy guidance and incentives, encouraging public data suppliers to open public data resources and promoting enterprise data suppliers to participate in data transactions can enrich the supply of the data factor market and enhance market activity.

Secondly, the differentiation and cultivation of intermediary service providers are key to the efficient operation of the data factor market. Intermediary service providers can include data processing service providers, data evaluation service providers, legal consulting service providers, and insurance service providers. Data processing service providers offer services such as data cleaning, labeling, and analysis to improve data quality. Data evaluation service providers assess data value, providing a reference for data pricing. Legal consulting service providers offer legal advice on data property rights and data security, safeguarding the legitimate rights and interests of trading parties. Insurance service providers offer data transaction insurance services to mitigate transaction risks. By cultivating specialized intermediary service organizations, the efficiency and security of data transactions can be enhanced, promoting the healthy development of the data factor market.

Thirdly, the differentiation and cultivation of demand-side parties are central to the demand side of the data factor market. Demand-side parties can include enterprise demanders, government demanders, and research institution demanders. Enterprise demanders primarily acquire data resources for product development, production and operations, and marketing to enhance core competitiveness. Government demanders acquire data resources for policy formulation, government services, and market regulation to improve public service efficiency and social governance capabilities. Research institution demanders acquire data resources for scientific research, such as academic studies and technological innovation. By cultivating diversified demand-side parties, the demand for the data factor market can be expanded, promoting its scaled development.

5. Evolution Mechanisms of Data Factor Markets

The maturation of data factor markets is driven by a closed-loop mechanism. External shocks alter the feasibility or necessity of data circulation. Internal feedback reflects strategic responses from suppliers, demanders, intermediaries, platforms, and regulators. Dynamic adjustment institutionalizes the lessons of these interactions through new standards, contracts, compliance tools, and pricing practices. This mechanism is non-linear because shocks do not automatically produce progress: they matter only when market participants and institutions absorb them. We provide a dynamic and evolutionary explanation of how data factor markets emerge, differentiate themselves from traditional factor markets, and gradually mature through the multifaceted interaction of technology, institutions, and strategic actors. Moving beyond conventional static analysis, this section's core contribution lies in conceptualizing data markets not as pre-configured institutional constructs, but as path-dependent, co-evolving systems shaped by the interplay of exogenous shocks—such as disruptive technological breakthroughs and endogenous strategic behaviors. By situating this evolution within the broader transition toward a digital economy, we argue that the data factor market represents a structural departure from the scarcity-driven logic of land and capital. Instead, it functions as a fluid ecosystem where the 'rules of the game' are iteratively refined.

5.1. Literature Development Stages and Characteristics

Drawing on domestic and international scholarship, policy developments, and market practice, we divided research on data factor markets (i.e., treating data as a factor of production) into three stages.

Nascent stage (pre-2015): Research attention in this stage centered on recognizing data's economic value and exploring early applications. As the digital economy was still emerging, the notion of data as a production factor had not yet been widely institutionalized (Cong et al., 2021). Studies therefore focused on how data could improve business decision-making, operational efficiency, and user experience in contexts such as the internet and e-commerce (Koutroumpis et al., 2020; J. Zhang et al., 2024). Although some scholars raised concerns about data security and privacy, the literature remained fragmented and lacked an integrated theoretical framework. These studies corroborate that, prior to 2015, data was viewed primarily as a tool for improving operational efficiency rather than being institutionally recognized and organized within a "factor market" framework as a factor of production.

Growth stage (2015–2020): Research attention in this stage shifted toward the institutional foundations and transaction arrangements of data markets. In China, a series of policy initiatives promoting market-oriented allocation of data accelerated interest in topics such as property rights definition, trading-platform design, and security and privacy regulation (Xue, 2025). Studies examined alternative approaches to allocating data rights, the operational mechanisms of data trading platforms, and complementary technical and regulatory safeguards, producing findings with direct implications for market design (Koutroumpis et al., 2020). Internationally, research also expanded into issues of cross-border data flows and governance, including compliance, regulatory coordination, and rule-setting (J. Zhang et al., 2024).

Mature stage (post-2025): Research attention in this stage diversified and deepened as markets and governance frameworks evolved. Research extended beyond foundational institutions to market dynamics, participant behavior, international comparisons, and global governance (Ye et al., 2022). Methodologically, scholars increasingly applied approaches such as evolutionary game models and econometric analysis to explain strategic interaction and the market's dynamic evolution (Ma et al., 2025; Sun et al., 2023). In parallel, advances in technologies such as privacy-preserving computation and large AI/foundation models prompted new studies on how technological innovation reshapes data market structure (Cong et al., 2021; Pathak, 2024; Driessen et al., 2022), valuation, and regulatory risk.

5.2. Mechanisms of Dynamic Evolution

5.2.1. External Catalysts: Technological Breakthroughs and Policy Shifts

Technological innovation serves as the exogenous engine driving the evolution of data factor markets. Notably, breakthroughs in technologies such as Large Language Models (LLMs), Privacy-Preserving Computing (PPC), and Quantum Computing have catalyzed significant shifts in how data is produced, circulated, and utilized. For example, LLMs have enabled multi-domain data integration, breaking down vertical silos in the data economy, particularly in industries such as healthcare, where they assist in predictive analytics and decision-making processes. Similarly, federated learning and secure multi-party computation have addressed privacy concerns, enabling data sharing across organizations without compromising individual data security. These technologies have lowered transaction barriers, expanded market access and created new avenues for data-driven innovations.

Policy recalibration plays a pivotal role in aligning incentives and regulating stakeholder behavior. For example, China's "Data 20" Guidelines and the EU Data Act have introduced frameworks that balance market autonomy with government oversight. The "Data 20" guidelines introduced a Tripartite Property Rights System, which has shifted data governance from centralized allocation to a more flexible hybrid market-government approach. Studies have shown that this shift has enhanced data flow across sectors like e-commerce and healthcare, fostering innovation while maintaining regulatory oversight.

The EU Data Act further strengthens market integration by establishing rules on fair access to and use of data across sectors, providing a legal framework that mitigates the tension between privacy rights and data utility. These policy interventions not only regulate but also incentivize market participants to align with broader economic and social objectives.

5.2.2. Internal Dynamics: Multi-Stakeholder Strategic Games

The data factor market is a dynamic, multifaceted ecosystem where governments, suppliers, intermediaries, and end-users engage in strategic interactions. For example, negotiations over data pricing, quality, and security are central to shaping the terms of data exchange. Negotiations between cloud service providers and enterprise clients over data storage fees and security provisions have become increasingly complex, as both parties seek to protect sensitive information while optimizing operational efficiency. Moreover, the relationship between state and market actors, especially the delicate balance between regulation and innovation, drives the continuous evolution of property rights and oversight mechanisms. As demonstrated by the ongoing regulatory adjustments in the U.S. and EU regarding data privacy laws, governments play an essential role in steering the market's development while encouraging technological innovation. The competition among data intermediaries to offer better services has further professionalized the market, ensuring greater transparency and efficiency in data exchanges.

5.2.3. Evolution of Price Discovery Mechanisms

The evolution of price discovery mechanisms is a crucial indicator of market maturity. In early pilot settings, pricing often relied on administrative guidance, cost-based heuristics, or negotiated arrangements rather than mature market signals, reflecting neither the true scarcity nor the utility of data. However, as market forces began to play a greater role, a hybrid model emerged, wherein the government still exerted influence through classification standards and price ceilings, but market negotiations became the primary mechanism for pricing. Today, as the market has matured, prices for data are determined through dynamic market interactions, where factors such as perceived value, market sentiment, and risk premiums come into play. This “autonomous” stage reflects the increasing sophistication of data markets, where prices reflect a more accurate allocation of data resources, akin to the Pareto-optimal conditions observed in well-functioning financial markets.

5.2.4. Practical Translation and Application Fields

The framework can be translated into practice in at least four fields. First, data exchanges can be used to diagnose whether weak trading activity results from poor data quality, unclear rights, inadequate compliance review, or insufficient demand scenarios. Second, public data operators can use the framework to design authorization, pricing, and accountability mechanisms that balance public value and commercial reuse. Third, industrial data spaces can apply the framework to coordinate suppliers, intermediaries, and demanders through common standards, APIs, and trusted computing environments. Fourth, competition and platform regulators can use the framework to distinguish efficiency-enhancing network effects from lock-in, exclusion, or monopoly risks. These applications clarify how a conceptual model can inform policy design and market governance.

6. Discussion and Conclusions

6.1. Discussion

This discussion underscores the theoretical necessity of placing the data factor market in explicit contrast with traditional factor markets—such as land, capital, and labor—rather

than treating it as their linear extension. Unlike these established markets, which are typically anchored in relatively stable property boundaries, standardized transaction rules, and mature institutional infrastructures, the data factor market entails a distinct economic logic shaped by data's non-rivalry, replicability, and strong externalities. Recognizing these structural and functional divergences is essential for accurately conceptualizing the data factor market as a pioneering, new-type factor market—one that is inherently “rule-intensive and infrastructure-intensive.” In this sense, its effective operation depends crucially on the synergistic resonance among technological support, institutional supply, and demand traction.

First, the economic attributes of data as a production factor dictate that its organizational logic must transcend traditional frameworks. Data possesses good quasi-public characteristics: high production/maintenance costs coupled with near-zero marginal costs for replication and reuse. Consequently, “tradability” does not arise naturally from scarcity but depends on institutional arrangements for controlled sharing and compliant utilization. Furthermore, the non-rivalry and excludability of data are “relative,” with excludability increasingly constructed through a blend of technology and regulation. In the absence of robust access control, compliance auditing, or liability definition, transaction risks escalate rapidly, stifling market vitality.

Additionally, the network effects and bilateral externalities of data imply that its value increases non-linearly with the number of participants, data scale, and cross-domain connectivity. While this drives platformization and ecologization, it also introduces governance challenges regarding market concentration and lock-in effects. Therefore, this study posits a composite theoretical framework—transaction costs, bundles of property rights, and network externalities—to explain market formation. The emergence of organized trading platforms is essentially a mechanism to reduce costs associated with information search, quality verification, contracting, and enforcement, thereby enhancing matching efficiency through scale expansion.

Second, the formation of the data factor market is driven by the synergy of technology, institutions, and economic demand. Technological advancement systematically reduces uncertainty by providing traceability, auditability, and value extraction capabilities. Institutional supply provides a stable behavioral framework through the demarcation of rights, compliance baselines, and standardized rules. Economic demand serves as the continuous traction force, as enterprises and the public sector seek data to enhance resource allocation and competitive advantage. These three elements form a functional chain: “Technology enables feasibility—Institutions enable legality—Demand enables scalability.”

Third, the evolutionary path of the data market exhibits a progression from “fragmented exchange” to “systematic organization” and finally to “cross-domain ecological synergy.” From 2015–2020 (Primary Stage): Characterized by pilot platforms and point-to-point exchanges of raw resources, constrained by ambiguous ownership and security risks. From 2021–2025 (Intermediate Stage): As property rights and rules clarified, participation surged. Transactions shifted from raw data to data products and services, supported by vertical markets and industry standards. Post-2025 (Advanced Stage): The focus shifts to cross-regional and cross-industry integration. A professionalized intermediary system (encompassing valuation, compliance, legal, insurance, and processing) becomes the backbone of market expansion and risk control.

At the dynamic mechanism level, we emphasize the coupling of external catalysis and internal gaming. Externally, breakthroughs in privacy computing and Large Language Models shift the “feasibility frontier” to a stepwise fashion. Internally, continuous bargaining among governments, suppliers, intermediaries, and demanders over price, quality, and liability drives the iteration of platform governance. Consequently, mature governance is

not merely a “design problem” but a “co-evolutionary problem” of structure and behavior, requiring a sustainable equilibrium between efficiency and competition, and openness and security.

Finally, the evolution of price discovery mechanisms serves as a proxy for market maturity. Prices evolve from administrative guidance or heuristic pricing toward “market negotiation under regulatory constraints,” and eventually toward value-based pricing that reflects specific scenarios and risk premiums.

From a policy perspective, the utility of the framework lies in identifying the binding constraint at each stage. When rights are unclear, market building should prioritize registration, authorization, and liability rules. When demand is weak, policy should support sectoral scenarios and data product standardization. When network effects become too strong, governance should focus on interoperability, portability, transparent access conditions, and competition safeguards.

6.2. Conclusions

This study demonstrates that the data factor market is not a spontaneous extension of ordinary commodity exchange, but an institutional market constructed through governance capacity, infrastructural support, and demand-side activation. Its formation and maturation depend on the coordinated interaction of technology, institutions, and demand: technological systems reduce transaction frictions and enable trusted circulation; institutional arrangements clarify rights, responsibilities, and compliance boundaries; and economic demand converts data from an internal organizational resource into a reallocable production factor. In this sense, the marketization of data is simultaneously a technical, institutional, and economic process.

The analysis shows that the distinctive attributes of data, especially its quasi-public good characteristics, relative non-rivalry, layered excludability, and strong network effects—fundamentally reshape the logic of factor-market construction. These attributes make data markets naturally inclined toward platform-based organization and ecosystem-based circulation. However, platformization does not automatically improve allocation efficiency. Its welfare effects depend on whether privacy protection, public-interest safeguards, interoperability rules, and competition governance can prevent data concentration, market lock-in, and asymmetric control over access.

In terms of formation mechanisms, this study identifies a technology–institution–demand synergy. Technological innovation provides the “trust layer” by supporting traceability, verification, privacy-preserving computation, and secure data use. Institutional supply defines the “legal boundary” by establishing rights delineation, compliance review, security classification, and dispute resolution mechanisms. Economic demand functions as the “growth engine” by linking data resources to concrete application scenarios, business models, public service needs, and research uses. Only when these three mechanisms reinforce one another can data become tradable, priceable, and governable.

Regarding evolutionary patterns, the data factor market follows a contingent pathway from government-led pilot allocation to vertical professionalization, and then to cross-domain ecosystem integration. This trajectory is not a linear or inevitable sequence, but a conditional process shaped by technological feasibility, rights clarity, demand liquidity, intermediary capacity, and regulatory adaptability. The price discovery mechanism evolves accordingly: early pricing often relies on administrative guidance, negotiated arrangements, or empirical valuation, while more mature stages require scenario-based, risk-adjusted, and market-responsive pricing models supported by professional intermediaries and standardized transaction rules.

The dynamic evolution of the data factor market is driven by the interaction between external shocks and internal strategic adjustment. Technological breakthroughs, policy reforms, and changes in application demand can expand market boundaries and create opportunities for leapfrogging development. At the same time, suppliers, demanders, intermediaries, platforms, and regulators continuously adjust their strategies, producing an iterative process of rule refinement, institutional learning, and governance upgrading. Network effects improve matching efficiency and value creation, but they also require dynamic regulatory tools to address over-concentration, interoperability barriers, and lock-in risks.

The theoretical contribution of this study lies in integrating transaction cost theory, the bundle-of-rights approach, and network externality theory into a unified framework for explaining data factor market formation and maturation. This framework clarifies why reducing transaction costs alone is insufficient: data markets can function effectively only when rights are operational, compliance is enforceable, intermediaries are professionalized, and platform power is subject to appropriate governance. High-quality market allocation therefore requires a balance between harnessing efficiency gains from network effects and mitigating concentration risks through proactive platform governance and competition policy.

Although the analysis is rooted in China's institutional context, the framework has broader relevance for other emerging data economies. It offers a diagnostic tool for regulators, data exchanges, platform firms, public data operators, and sectoral data-space builders to identify the binding constraint in market development—whether it lies in technical feasibility, rights clarity, demand liquidity, intermediary capacity, or competition risk. Future empirical research can test the framework by linking rights registration reforms, transaction records, data product listings, price information, network concentration indicators, and regional policy variation to observable changes in data circulation, market performance, and institutional maturity.

Funding: This research was funded by the National Natural Science Foundation of China (No. 42371269).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: There is no new data being created.

Acknowledgments: During the preparation of this manuscript, the author used Claude (Opus 4.7, Anthropic) for grammar checking and language editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

- Abrardi, L., Cambini, C., Congiu, R., & Pino, F. (2024a). User data and endogenous entry in online markets. *Journal of Industrial Economics*, 72(3), 1052–1088. [\[CrossRef\]](#)
- Abrardi, L., Cambini, C., & Pino, F. (2024b). Regulating data sales: The role of data selling mechanisms. *Telecommunications Policy*, 48(8), 102813.
- Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Review*, 112(3), 887–925.
- Anand, V., Magesh, G., Alagiri, I., Brahmam, M., Balusamy, B., Selvan, C., Alshahrani, H., Getahun, M., & Soufiene, B. (2025). Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions. *Scientific Reports*, 15, 1615. [\[CrossRef\]](#)
- Azcoitia, S., & Laoutaris, N. (2022). A survey of data marketplaces and their business models. *ACM SIGMOD Record*, 51, 18–29. [\[CrossRef\]](#)

- Bajari, P., & Tadelis, S. (2011). Incentives versus transaction costs: A theory of procurement contracts. *RAND Journal of Economics*, 42(3), 387–407.
- Belleflamme, P., & Peitz, M. (2019). *Industrial organization: Markets and strategies* (2nd ed.). Cambridge University Press.
- Brynjolfsson, E., Rock, D., & Syverson, C. (2018). Artificial intelligence and the modern productivity paradox. *Journal of Economic Perspectives*, 32(3), 3–30.
- Chen, H. (2023). Enterprise marketing strategy using big data mining technology combined with XGBoost model in the new economic era. *PLoS ONE*, 18(6), e0285506. [CrossRef] [PubMed]
- Chen, J., Li, X., & Luo, X. (2021). Data trading, platform governance, and market efficiency. *Information Systems Research*, 32(4), 1208–1227.
- Chen, L. (2023). A study on the economic drivers of the data element. *Advances in economics. Management and Political Sciences*, 4, 358–365. [CrossRef]
- Chen, R. (2023). The legal structure of structural separation of data property rights: Chinese theory of data property rights. *US-China Law Review*, 20(7), 315–321. [CrossRef]
- Chen, Y., Liu, M., Zhang, J., Tan, A., & Wang, Y. (2024). A three-party dynamic pricing mechanism for customized data products based on the stackelberg game and bargaining model. *Symmetry*, 16, 1540. [CrossRef]
- Cong, L., Xie, D., & Zhang, L. (2021). Knowledge accumulation, privacy, and growth in a data economy. *LSN: Other Regulation of Information & Privacy Issues Involving Consumers (Sub-Topic)*, 9, 10028.
- CPC Central Committee & State Council of the People's Republic of China. (2022, December 19). *Opinions on building a basic data system to better tap the role of data as a factor of production*. Available online: http://www.gov.cn/zhengce/2022-12/19/content_5732085.htm (accessed on 29 May 2026).
- Creemers, R. (2022). China's emerging data protection framework. *Journal of Cybersecurity*, 8, tyac011. [CrossRef]
- Crémer, J., de Montjoye, Y. A., & Schweitzer, H. (2019). *Competition policy for the digital era*. European Commission Report. European Commission.
- Deepa, N., Pham, V., Nguyen, D., Bhattacharya, S., Prabadevi, B., Gadekallu, T., Maddikunta, P., Fang, F., & Pathirana, P. (2022). A survey on blockchain for big data: Approaches, opportunities, and future directions. *Future Generation Computer Systems*, 131, 209–226. [CrossRef]
- Driessen, S., Monsieur, G., & Heuvel, W. (2022). Data market design: A systematic literature review. *IEEE Access*, 10, 33123–33153. [CrossRef]
- Duan, H., Wang, L., Wu, T., & Wang, J. (2025). Research on the marketization of data elements empowering the digital development of industrial economy. *International Journal of Finance and Investment*, 1(1), 16–21. [CrossRef]
- Duch-Brown, N., Martens, B., & Mueller-Langer, F. (2017). The economics of ownership, access and trade in digital data. *JRC Working Papers on Digital Economy*. [CrossRef]
- Eichler, R., Gröger, C., Hoos, E., Schwarz, H., & Mitschang, B. (2022). From data asset to data product—The role of the data provider in the enterprise data marketplace. In J. Barzen, F. Leymann, & S. Dustdar (Eds.), *Service-oriented computing. SummerSOC 2022. Communications in computer and information science* (Vol. 1603). Springer. [CrossRef]
- European Commission. (2020). *A European strategy for data*. European Commission.
- European Union. (2016). *General data protection regulation (EU) 2016/679*. European Union.
- Fast, V., Schnurr, D., & Wohlfarth, M. (2023). Regulation of data-driven market power in the digital economy: Business value creation and competitive advantages from big data. *Journal of Information Technology*, 38, 202–229. [CrossRef]
- Fernandez, R., Subramaniam, P., & Franklin, M. (2020). Data market platforms: Trading data assets to solve data problems. *Proceedings of the PVLDB Endowment*, 13(11), 1933–1947. [CrossRef]
- Gao, X., & Shao, Z. (2021). Legal boundaries of data property rights in the digital economy. *Computer Law & Security Review*, 41, 105528.
- Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1), 3–43. [CrossRef]
- Guo, X., Zhang, X., Yang, J., Wei, J., Tamamine, T., Lin, J., & Hu, W. (2025). Digital transformation, entrepreneurship, and total factor productivity of enterprises: Empirical evidence from listed companies. *International Entrepreneurship and Management Journal*, 21, 102. [CrossRef]
- Hernández-Tamurejo, Á., Lacárcel, F. J. S., & Corvello, V. (2025). The economics of AI adoption in OTAs: Market dynamics and future research. *Global Economics Research*, 1(1), 100001. [CrossRef]
- Hoofnagle, C., van der Sloot, B., & Zuiderveen Borgesius, F. (2019). The European union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, 28(1), 65–98. [CrossRef]
- Huang, J. (2024). The rise of data property rights in China: How does it compare with the EU data act and what does it mean for digital trade with China? *Journal of International Economic Law*, 27(3), 462–479. [CrossRef]
- Huang, L., Dou, Y., Liu, Y., Wang, J., Chen, G., Zhang, X., & Wang, R. (2021). Toward a research framework to conceptualize data as a factor of production: The data marketplace perspective. *Fundamental Research*, 1(5), 586–594. [CrossRef]

- Jarke, M., Otto, B., & Ram, S. (2019). Data sovereignty and data space ecosystems. *Business & Information Systems Engineering*, 61, 549–550. [CrossRef]
- Juárez-Varón, D., & Monreal, E. G. (2025). Market-oriented entrepreneurship and the impact of social media and crowdsourcing on continuous organizational learning: An empirical study. *Sustainable Technology and Entrepreneurship*, 4(1), 100088. [CrossRef]
- Ke, T., & Sudhir, K. (2022). Privacy rights and data security: GDPR and personal data markets. *Management Science*, 69, 4389–4412. [CrossRef]
- Koutroumpis, P., Leiponen, A., & Thomas, L. (2020). Markets for data. *Industrial and Corporate Change*, 29, 645–660. [CrossRef]
- Li, J., Li, J., Wang, X., Qin, R., Yuan, Y., & Wang, F. (2023). Multi-blockchain based data trading markets with novel pricing mechanisms. *IEEE/CAA Journal of Automatica Sinica*, 10, 2222–2232. [CrossRef]
- Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading and protection. *IEEE Access*, 6, 15132–15154. [CrossRef]
- Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Transactions on Industrial Informatics*, 16, 4177–4186.
- Ma, J., Yu, J., & Cheng, Y. (2025). Dynamic evolution analysis of incentive strategies and symmetry enhancement in the personal-data valorization industry chain. *Symmetry*, 17, 1639. [CrossRef]
- Magas, M., & Kiritsis, D. (2021). Industry commons: An ecosystem approach to horizontal enablers for sustainable cross-domain industrial innovation (a positioning paper). *International Journal of Production Research*, 60, 479–492. [CrossRef]
- Margoni, T., Ducuing, C., & Schirru, L. (2023). Data property, data governance and common European data spaces. *Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht*. [CrossRef]
- Martens, B. (2024). *An institutional economics approach to data spaces as data market intermediaries*. Available at SSRN. Available online: <https://ssrn.com/abstract=4731126> (accessed on 22 June 2026).
- Möller, F., Jussen, I., Springer, V., Gieß, A., Schweihoff, J., Gelhaar, J., Guggenberger, T., & Otto, B. (2024). Industrial data ecosystems and data spaces. *Electronic Markets*, 34, 41. [CrossRef]
- OECD. (2019). *Enhancing access to and sharing of data*. OECD Publishing.
- OECD. (2021). *Data governance and competition policy in the digital age*. OECD Publishing.
- Parker, G., Van Alstyne, M., & Choudary, S. (2016). *Platform revolution*. W. W. Norton.
- Pathak, M. (2024). Data governance redefined. *European Data Protection Law Review*, 10(1), 43–56. [CrossRef]
- Posner, E. A., & Weyl, E. G. (2017). Property is only another name for monopoly. *Journal of Legal Analysis*, 9(1), 51–123. [CrossRef]
- Rochet, J. C., & Tirole, J. (2006). Two-sided markets: A progress report. *RAND Journal of Economics*, 37(3), 645–667. [CrossRef]
- Schweihoff, J., Lipovetskaja, A., Jussen-Lengersdorf, I., & Möller, F. (2024). Stuck in the middle with you: Conceptualizing data intermediaries and data intermediation services. *Electronic Markets*, 34, 48. [CrossRef]
- Sun, Y., Zhang, Y., Li, J., & Zhang, S. (2023). Evolutionary game analysis of data resale governance in data trading. *Systems*, 11(7), 363. [CrossRef]
- Trakadas, P., Masip-Bruin, X., Facca, F., Spantideas, S., Giannopoulos, A., Kapsalis, N., Martins, R., Bosani, E., Molinas-Ramón, J., Prats, R., Ntoulas, G., & Lyridis, D. (2022). A reference architecture for cloud-edge meta-operating systems enabling cross-domain, data-intensive, ML-assisted applications: Architectural overview and key concepts. *Sensors*, 22, 9003. [CrossRef] [PubMed]
- Wang, D., & Yang, T. (2025). Research on the promotion effect of the marketization of data elements on the digital transformation of manufacturing enterprises: An empirical evaluation of a multiperiod DID model. *Sustainability*, 17, 3199. [CrossRef]
- Wang, H. (2025). Mechanism innovation and policy optimization for market-based allocation of data elements in the digital economy era. *Frontiers in Economics & Policy Modeling*, 1(1), 22–26. [CrossRef]
- Williamson, O. E. (2010). Transaction cost economics: The natural progression. *American Economic Review*, 100(3), 673–690. [CrossRef]
- Xu, J., Hong, N., Xu, Z., Zhao, Z., Wu, C., Kuang, K., Wang, J., Zhu, M., Zhou, J., Ren, K., Yang, X., Lu, C., Pei, J., & Shum, H. (2023). Data-driven learning for data rights, data pricing, and privacy computing. *Engineering*, 25, 66–76. [CrossRef]
- Xue, Y. (2025). Research on the data elements enabling new quality productivity mechanisms, challenges, and pathways. *SHS Web of Conferences*, 218, 04011. [CrossRef]
- Yang, B. (2025). Data element marketization and energy efficiency of heavy polluting enterprises: A technology innovation perspective. *Journal of Environmental Management*, 391, 126318. [CrossRef] [PubMed]
- Ye, Y., Zhou, A., Shi, X., & Huang, C. (2022). A SEED model for constructing the data factor market: Evidence from Guiyang global big data exchange (GBDEX) in China. *Journal of Digital Economy*, 1(3), 273–283. [CrossRef]
- Yu, Y., Yu, J., Wang, X., Li, J., Lin, Y., He, C., Yang, Y., Qiao, Y., Li, L., & Wang, F. (2024). Navigating the data trading crossroads: An interdisciplinary survey. *arXiv*, arXiv:2407.11466.
- Yuan, H., & Du, X. (2025). Exploring the mechanism of how the market-based allocation of data elements affects the supply chain resilience of manufacturing enterprises: A perspective on data as a production factor. *Sustainability*, 17, 7950. [CrossRef]

- Zhang, J., Bi, Y., Cheng, M., Liu, J., Ren, K., Sun, Q., Wu, Y., Cao, Y., Fernandez, R., Xu, H., Jia, R., Kwon, Y., Pei, J., Wang, J., Xia, H., Xiong, L., Yu, X., & Zou, J. (2024). A survey on data markets. *arXiv*, arXiv:2411.07267.
- Zhang, M., Beltrán, F., & Liu, J. (2023). A survey of data pricing for data marketplaces. *IEEE Transactions on Big Data*, 9, 1038–1056. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.