

Data Elements as a New Factor of Production: A Critical and Integrative Literature Review

Yu Shen

Nanjing University of Science and Technology ZiJin College, Nanjing 210023, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Data has rapidly evolved from a by-product of digital activities into an economically consequential resource that can be collected, processed, reused, exchanged, and embedded in organizational decision-making. This evolution has stimulated a growing literature across economics, management, information systems, public policy, law, and accounting, but the literature remains conceptually fragmented. In particular, scholars use the terms data resource, data asset, data element, and data factor to describe overlapping but not identical phenomena. This paper develops a critical and integrative review of the literature on data as a production factor, with particular attention to the institutionalization and marketization of data elements in China. The review is organized around four connected questions: (1) what economic characteristics distinguish data from conventional production factors; (2) how data can be transformed from a resource into a market-allocable factor; (3) how property rights, privacy, security, standards, and governance shape data circulation; and (4) through what mechanisms data-factor marketization affects productivity, innovation, digital transformation, and sustainable development. International research emphasizes the non-rival nature of data, data externalities, privacy, information asymmetry, and market design. China-focused research places greater emphasis on factor recognition, data trading institutions, market construction, property-right arrangements, and empirical identification using data-trading-platform establishment as a quasi-natural experiment. Recent evidence links data-factor marketization to enterprise digital transformation, innovation quality, regional innovation, green innovation, land-use efficiency, and other economic outcomes. However, major gaps remain in measurement and valuation, causal identification, heterogeneous data quality, cross-organizational spillovers, and the interaction between market mechanisms and governance. On this basis, the paper proposes an integrated framework in which data characteristics form the technological foundation, institutional arrangements enable circulation, market mechanisms allocate data, complementary capabilities convert data into productive inputs, and economic and social outcomes emerge from these interactions. The review concludes with a research agenda emphasizing interdisciplinary theory building, stronger measurement, causal inference, and comparative institutional analysis.

Keywords: Data elements; Data factor; Data factor marketization; Data market; Data governance; Digital transformation; Innovation; Literature review

Online publication: September 30, 2026

1. Introduction

The rapid development of cloud computing, artificial intelligence, digital platforms, Internet of Things technologies, and data-intensive business models has changed the economic role of information. In traditional economic analysis, information is often treated as an input that reduces uncertainty or improves decision-making. In contemporary digital production systems, however, data can itself become an object of investment, processing, exchange, and strategic control. A firm's customer records, transaction histories, sensor streams, logistics data, or training datasets can be repeatedly reused, combined with other datasets, and fed into algorithms. The resulting economic value does not necessarily arise from the data alone; it emerges from the interaction between data and complementary technologies, human capital, organizational capabilities, and institutions ^[1-4].

The economics literature has provided an important theoretical foundation for understanding this transformation. Goldfarb and Tucker ^[1] showed that digital technologies reduce several categories of economic friction, including search, replication, transportation, tracking, and verification costs. Jones and Tonetti ^[2] emphasized a distinctive characteristic of data: non-rivalry. The same data can be used by multiple firms without being physically depleted, creating the possibility of increasing returns and broad social spillovers. Bergemann, Bonatti, and Gan ^[3] further demonstrated that individual data can generate social externalities because one person's information may predict the behavior of others. At the same time, Acemoglu *et al.* ^[4] showed that data markets can generate inefficient outcomes because information externalities can depress prices and induce excessive data sharing. These studies collectively suggest that the economics of data cannot be understood simply by applying conventional commodity-market assumptions.

China provides an especially important institutional context for studying this transformation. Data was formally recognized as a factor of production in national factor-allocation policy, and subsequent policy initiatives have focused on data trading platforms, data-rights arrangements, data registration, public-data authorization, data-service providers, and the construction of a unified data-element market. The resulting literature differs in emphasis from much of the international literature. International work has generally developed from digital economics, privacy economics, industrial organization, and information economics, whereas Chinese research has increasingly addressed how data becomes an institutionalized production factor and how markets can be built to facilitate its circulation. Zhang *et al.* ^[5], using bibliometric methods on 1,304 international publications and 1,249 Chinese publications, identified substantial differences in governance environment, governance actors, governance technologies, and governance mechanisms between the two literatures.

The rapid expansion of empirical studies makes a new synthesis necessary. Recent work has moved beyond conceptual discussions to examine whether data-factor marketization changes enterprise digital transformation ^[6], innovation quality ^[7], regional innovation ^[8], green innovation ^[9], urban green development ^[10], and industrial land-use efficiency ^[11]. This development is important because it shifts the research question from "Is data a factor?" to "Under what institutional and organizational conditions does data become a productive factor, and through what mechanisms does its market allocation create value?"

As shown in **Figure 1**, this review addresses that question by integrating four streams of research: the economic characteristics of data; the institutionalization and marketization of data factors; governance, privacy, and property rights; and economic consequences of data-factor marketization. Rather than providing a chronological bibliography, the paper critically compares theoretical arguments, empirical strategies, and

unresolved issues. The central proposition is that data does not automatically become productive merely because it is abundant. Its economic role depends on an institutional process that makes data identifiable, usable, shareable, protectable, and allocable, and on complementary capabilities that transform data access into productive activity.

Figure 1. Evolutionary Logic of Data as an Economic Factor

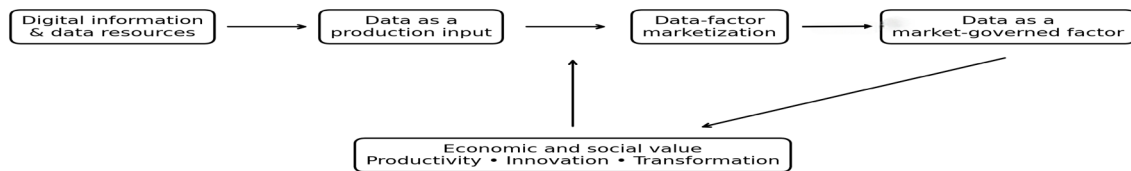


Figure 1. Evolutionary logic of data as an economic factor

2. Review scope and methodological approach

This paper adopts an integrative literature-review approach rather than a formal meta-analysis. The objective is to connect theoretical, institutional, and empirical research that uses partially different vocabularies to study a common phenomenon. An integrative review is appropriate because the field combines economics, management, law, information systems, public policy, and accounting, and because many important contributions are conceptual or institutional rather than statistically comparable.

Literature identification focused on peer-reviewed journal articles, authoritative working papers where the published version was not yet available, and recent review studies addressing data economics, data markets, privacy, data governance, data-factor marketization, and economic outcomes. International sources were prioritized from established economics and interdisciplinary journals, while China-focused research was included when it provided direct evidence on data-factor institutions or marketization. Recent studies published in 2024–2026 were given particular attention because the institutionalization of data-factor markets is developing rapidly. Bibliographic details and DOI information were checked against publisher, society, journal, or recognized bibliographic records before inclusion.

The review is organized around five analytical categories: (1) conceptual foundations and economic characteristics; (2) assetization and marketization; (3) governance, rights, privacy, and market design; (4) economic and organizational consequences; and (5) methodological and theoretical gaps. This organization allows the literature to be evaluated according to both what it has established and what remains unresolved.

3. Conceptual foundations: Why can data be a production factor?

3.1. Data and the economics of digital production

One starting point is the broader transformation of economic activity under digital technologies. Goldfarb and Tucker ^[1] conceptualized digital economics through the reduction of several economic frictions. Digitalization lowers the cost of searching for information, copying and distributing digital products, tracking behavior, and

verifying transactions. These changes alter the organization of markets and firms because information can move faster and at lower cost. From this perspective, the economic importance of data is initially derived from its role in reducing uncertainty and improving coordination.

The next conceptual step is to recognize that data can itself be accumulated and repeatedly used as an input. This distinguishes data from a one-time information service. A dataset may support multiple analytical tasks, be combined with other datasets, and generate new information through algorithms. Jones and Tonetti^[2] formalized this characteristic as non-rivalry. If data is non-rival, one firm's use does not necessarily reduce another firm's ability to use the same underlying information. This creates a potential source of increasing returns that is unusual among traditional production factors. Evidence from U.S. manufacturing also indicates that data-driven decision-making is associated with higher value-added and that organizational practices can be important complements to data use^[20].

Non-rivalry also creates a policy paradox. From a social perspective, broad use of non-rival data may create substantial benefits, but firms may have incentives to restrict access because sharing data can reduce competitive advantages. Jones and Tonetti^[2] therefore argued that data hoarding may produce inefficiently low social use. The implication for data-factor markets is that market design cannot focus only on ownership and price; it must also consider access, interoperability, competition, and incentives for sharing.

3.2. Data externalities and complementarity

Another distinctive property is that the value of one person's or organization's data may depend on information generated by others. Bergemann *et al.*^[3] showed that individual data can have a social dimension because information about one consumer can predict the behavior of other consumers. Such externalities complicate the idea that data can be valued independently. A dataset can create value beyond the direct transaction because it improves predictions, matching, targeting, or learning for other actors.

Acemoglu *et al.*^[4] provided a complementary warning. When data is interconnected, sharing by one individual can reveal information about others. This externality can reduce the marginal value of privacy protection and lead to prices that do not reflect the full social cost of data sharing. Their model shows that competitive data markets do not necessarily solve the problem. Thus, the existence of a market does not automatically guarantee efficient data allocation.

These insights imply that data should be treated as a relational rather than purely individual factor. Its value depends on interactions among datasets, users, firms, technologies, and institutions. This also helps explain why data markets often resemble markets for intermediate inputs and information services rather than conventional commodity markets.

3.3. From data resources to data assets and data factors

The literature increasingly distinguishes between raw data and economically organized data. A raw record may have little immediate value because it is incomplete, noisy, poorly structured, or legally unusable. A data resource is a collection that has potential productive usefulness. A data asset is a resource that has been organized, controlled, processed, and connected to a recognized economic purpose. A data factor is a broader institutional-economic concept: it emphasizes the role of data in production and resource allocation, including the institutions that make its circulation possible.

This distinction matters because "data" is not a homogeneous economic object. Quality, timeliness,

uniqueness, accuracy, coverage, interoperability, legality, and application specificity can radically change value. A highly specific dataset may be extremely valuable to one firm but almost useless to another. Conversely, a standardized dataset may have modest value in one application but large network value when combined with many other datasets. Consequently, a mature data-factor theory must move beyond a simple quantity-based view of data.

4. Marketization: From data resource to allocatable factor

4.1. Why data markets are institutionally difficult

A conventional market assumes that the object being exchanged can be identified, transferred, and priced with reasonable clarity. Data challenges all three assumptions. First, a buyer may acquire a right to access or use data rather than exclusive possession. Second, the seller may retain the ability to use the same data. Third, data quality and usefulness are often difficult to evaluate before application. Fourth, the value of a dataset depends on complementary assets held by the buyer. Fifth, privacy and security risks can create costs that are difficult to internalize in a bilateral transaction.

These characteristics help explain why data markets often require intermediaries, technical standards, contracts, security mechanisms, and trusted platforms. The market is therefore not simply a place where datasets are bought and sold. It is an institutional system that reduces uncertainty around data rights, quality, access, compliance, and use.

4.2. Data marketplaces and transaction costs

Huang *et al.* ^[12] developed a research framework for conceptualizing data as a factor of production from the data-marketplace perspective. Their framework highlights the importance of data lifecycle maturity and asset specificity. The argument is useful because it connects data-market development to transaction-cost economics: as data becomes more standardized and reusable, the cost of matching buyers and sellers can decline; where data is highly specific or immature, bilateral arrangements may remain more efficient.

From this perspective, a data exchange performs several functions simultaneously. It can provide a marketplace, establish technical and contractual standards, facilitate identity and compliance checks, and reduce search and negotiation costs. It may also create reputational mechanisms that improve trust. The development of data-factor markets therefore depends on complementary infrastructure rather than on trading demand alone.

4.3. China's institutionalization of data-factor markets

China represents a distinctive case because the recognition of data as a production factor has been accompanied by active institutional construction. Data trading platforms have become observable institutional milestones that researchers can use to identify marketization. Zhang *et al.* ^[5] showed that Chinese research has paid particular attention to data factors, enterprise data, transaction costs, data assets, data circulation, and data security, while international literature has relatively stronger emphasis on privacy technologies, smart contracts, and platform governance.

The difference should not be interpreted as a simple opposition between “market” and “governance.” Instead, it reflects different research stages. Where the fundamental challenge is to establish a market institution, scholarship naturally focuses on rights, platforms, transactions, and factor allocation. Where digital markets are already mature, research may focus more on competition, privacy, algorithmic governance, and data

externalities.

A recent MDPI study on the structural transformation of China's data-factor markets similarly argues that the institutional logic of tradability, priceability, and governability remains under-theorized and proposes a coevolutionary perspective linking factor market theory, platform theory, digital political economy, and comparative institutional analysis ^[13]. This line of research supports the argument that data-factor marketization should be analyzed as institutional transformation rather than simply as technological modernization. **Figure 2** illustrates this mechanism.

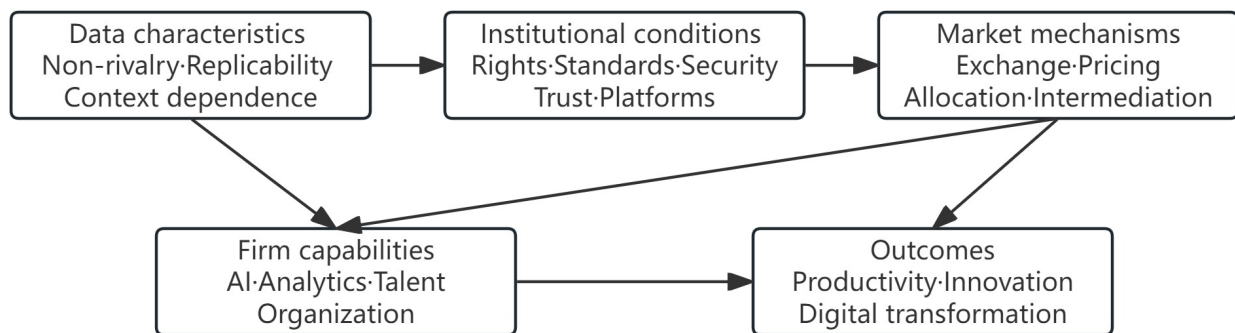


Figure 2. Integrated mechanism linking data factors to value creation

5. Governance, rights, privacy, and market design

5.1. Why property rights are necessary but insufficient

Property rights are central to factor markets because they define who can use, exclude, transfer, and benefit from a resource. However, data does not fit neatly into the conventional model of exclusive ownership. Jones and Tonetti ^[2] showed that consumer property rights can improve allocation by giving individuals greater control over data use, but the optimal design must balance privacy against the social benefits of broad data utilization. Acemoglu *et al.* ^[4] further showed that data externalities may make market prices socially misleading.

These findings suggest that data governance may need a bundle-of-rights approach. Different actors may hold rights to access, process, use, derive value from, or authorize particular applications of data without any single actor possessing an absolute ownership right. Such a framework is particularly relevant to enterprise and public data, where generation, collection, processing, and use may involve multiple organizations.

5.2. Privacy and security as economic variables

Privacy is not simply a legal constraint on data markets; it is also an economic variable that affects willingness to share, consumer behavior, firm strategy, and social welfare. Acquisti *et al.* ^[14] reviewed the economics of privacy and emphasize that privacy decisions are complicated by information asymmetry: individuals may not fully understand when data is collected, how it is used, or what consequences follow from disclosure. This asymmetry means that market mechanisms alone may not adequately protect social welfare.

Goldfarb and Tucker's edited volume on the economics of privacy further emphasizes open questions concerning the value of privacy, property rights, markets, inequality, and regulation ^[15]. For data-factor

markets, the implication is that privacy protection should be integrated into market design. Privacy-preserving computation, data minimization, controlled access, and secure data spaces can potentially reduce the trade-off between data use and privacy.

5.3. Governance as market infrastructure

A useful conceptual shift is to regard governance as market infrastructure. Standards determine whether data can be combined; security rules determine whether data can circulate safely; rights arrangements determine who can authorize use; and enforcement mechanisms determine whether contracts are credible. Without these institutions, even high demand for data may not translate into efficient exchange.

As shown in **Table 1**, the bibliometric review by Zhang *et al.* ^[5] is particularly relevant here. Its comparison of international and Chinese research shows that governance is not a single dimension. It involves the governance environment, actors, technologies, and mechanisms. This suggests that future research should avoid using a single index of “data governance” and instead examine how specific institutions affect specific market outcomes.

Table 1. Governance as market infrastructure

Governance dimension	Core question	Potential mechanism	Research implication
Rights	Who can access, use, authorize, and benefit?	Reduces uncertainty and disputes	Study bundles of rights rather than ownership alone
Quality & standards	Can datasets be compared and combined?	Reduces information and integration costs	Develop measurable quality and interoperability indicators
Privacy & security	How can data be used without unacceptable harm?	Changes sharing incentives and transaction costs	Model privacy as an economic variable
Platforms & contracts	How are exchanges organized?	Reduces search, matching, and enforcement costs	Study platform design and intermediary roles
Competition	Does market power restrict data access?	Affects prices, innovation, and sharing	Connect data markets with industrial organization

6. Economic and organizational consequences of data-factor marketization

6.1. Digital transformation

One of the most visible empirical outcomes is enterprise digital transformation. Zhan *et al.* ^[6] used the construction of data trading platforms in China as a quasi-natural experiment and report that data-element marketization significantly promotes enterprise digital transformation. Their mechanism analysis identifies digital technology innovation, digital talent attraction, and fintech development as important channels, with effects particularly visible in cloud computing, big data, and artificial intelligence technologies.

This finding supports a complementarity interpretation. Data-factor marketization does not simply supply firms with additional datasets. It can alter the incentives and conditions under which firms acquire technology, recruit talent, and develop digital capabilities. In other words, marketization can change the firm’s production environment, not merely its input quantity.

6.2. Innovation quantity and quality

The innovation literature is developing rapidly. Xu *et al.* ^[7] used the staggered implementation of data trading platforms across Chinese cities and patent citations to examine innovation quality. They find that data-factor

marketization significantly improves enterprise innovation quality and identify information infrastructure, financing constraints, supplier relationships, and intangible assets as mechanisms. The result is important because it moves beyond patent counts and suggests that data institutions may affect the quality and economic relevance of innovation.

Ouyang and Hu ^[8] provided complementary regional evidence using provincial panel data from 2000–2023. Their study finds that data factor marketization promotes regional innovation, with technological iteration and research-achievement transformation as key mechanisms. The effects are stronger in regions with higher economic development, marketization, R&D investment, and baseline innovation capacity. This heterogeneity is theoretically important: the value of a new factor depends on the capacity of the system receiving it.

6.3. Green innovation and sustainable development

The sustainability literature provides another important dimension. Song and Xu ^[9] reported that data-element marketization promotes corporate green innovation, with financing constraints, human-capital structure, and collaborative innovation identified as mechanisms. Li *et al.* ^[10] examined urban green development and showed that data-factor marketization is connected with reduced urban environmental pollution through digital transformation mechanisms. Chen and Li ^[11] similarly found that data-factor marketization improves urban industrial land-use efficiency, with technological innovation and reduced land-resource misallocation as important mechanisms.

These studies suggest that the economic effects of data factors extend beyond conventional productivity indicators. Data can facilitate coordination, improve matching, reduce resource misallocation, and accelerate technological upgrading. However, caution is necessary when interpreting these findings. Most empirical studies use the establishment of data trading platforms as the key policy shock. Platform establishment captures institutional change, but it may also correlate with local digital infrastructure, government capacity, or broader development strategies. Thus, causal interpretation requires careful identification and robustness analysis.

6.4. Productivity and new-quality productive forces

Recent Chinese research has connected data-factor marketization with new-quality productive forces. Dong *et al.* ^[16] reviewed the literature from the perspective of new-quality productive forces and argue that data's role has evolved through information-resource carrying, value-form transformation, and factor integration. They organize the mechanism through a “factor–technology–organization–pattern” framework and emphasize that data's effects are multidimensional: productivity enhancement may coexist with digital paradoxes, industrial restructuring may coexist with market imbalance, and aggregate growth may coexist with regional inequality.

This perspective helps broaden the literature beyond a simple production-function approach. Data is not merely another unit of input. It can change how production is organized, how technologies are selected, how markets are coordinated, and how knowledge circulates. The economic effect of data therefore depends on systemic complementarity.

7. Synthesis of the literature

The literature on data as a factor of production has developed rapidly across economics, information systems, management, and public policy. Although existing studies differ in their research questions and empirical settings, they can be broadly organized into six interconnected streams: data economics, data marketplaces,

governance and privacy, enterprise effects, regional effects, and the institutionalization of data as a factor of production. These streams represent different dimensions of the same underlying process through which data is transformed from a digital resource into an economically productive factor.

The first stream focuses on the economic characteristics of data and provides the theoretical foundation for treating data as a distinctive input in production and economic activity. Goldfarb and Tucker ^[1] emphasized that digital technologies can reduce information, search, communication, and transaction costs, thereby changing the organization of economic activity. Jones and Tonetti ^[2] further highlighted the non-rival nature of data. Unlike traditional physical inputs, the same dataset can potentially be used by multiple agents without being depleted. This characteristic creates possibilities for extensive reuse and knowledge spillovers, but it also generates challenges for conventional market mechanisms. Bergemann *et al.* ^[3] and Acemoglu *et al.* ^[4] extended this discussion by examining information, pricing, and efficiency in data markets. Together, these studies establish an important microeconomic foundation: the economic value of data does not arise simply from its existence, but from its potential to reduce uncertainty, improve decision-making, and generate additional value when combined with other productive resources.

The second stream examines data marketplaces and the institutional mechanisms through which data can be exchanged. Huang *et al.* ^[12] conceptualized data from a marketplace perspective and emphasize its lifecycle, specificity, and the institutional arrangements required for transactions. Unlike standardized physical commodities, data is often heterogeneous, context-dependent, difficult to evaluate, and potentially sensitive. These characteristics increase transaction costs and make conventional market mechanisms insufficient. Therefore, the development of data markets requires intermediaries, standards, pricing mechanisms, quality assessment, and mechanisms for establishing trust. This literature is particularly important because it shifts the discussion from the abstract question of whether data has economic value to the more practical question of how such value can be realized through market exchange. However, existing research still provides relatively limited evidence concerning the long-term performance, efficiency, and stability of emerging data markets.

The third stream concerns data governance, privacy, security, and institutional legitimacy. The expansion of data circulation creates a fundamental tension between economic utilization and the protection of individual and organizational rights. Acquisti *et al.* ^[14] demonstrated that privacy has important economic dimensions and cannot be considered solely as a legal or ethical issue. Similarly, Goldfarb and Tucker ^[15] emphasized that digital markets are strongly affected by institutional and technological conditions. Recent research on data-factor institutions has further examined how rights, trading arrangements, and governance mechanisms shape the development of data markets ^[5]. These studies suggest that data marketization and data governance are not independent processes. Instead, effective marketization requires clearly defined rights, appropriate privacy protection, data security, and institutional trust. Nevertheless, questions concerning the specific composition of data rights, cross-jurisdictional institutional differences, and the balance between data circulation and privacy protection remain insufficiently explored.

The fourth stream investigates the effects of data-factor marketization at the enterprise level. This literature moves beyond the conceptual and institutional dimensions of data and examines whether data can generate observable economic outcomes within firms. Zhan *et al.* ^[6] found that data-factor marketization is associated with enterprise digital transformation, while Xu *et al.* ^[7] examined its relationship with the quality of enterprise innovation. These studies indicate that access to data alone does not automatically produce economic value. Rather, firms need complementary capabilities, including digital infrastructure, organizational flexibility,

technological capabilities, and managerial expertise, to convert data into productive outcomes. From this perspective, data should not be viewed as an isolated input. Its economic contribution depends on its interaction with other factors of production, particularly technology, human capital, organizational capabilities, and knowledge.

The fifth stream extends the analysis from firms to regional and spatial economic effects. Recent studies have examined the relationship between data-factor marketization and regional innovation, green development, and resource-allocation efficiency. Ouyang and Hu ^[8] investigated the relationship between data-factor marketization and regional innovation, while Song and Xu ^[9] examined its implications for corporate green innovation. Li *et al.* ^[10] connected digital transformation with urban green development, and Chen and Li ^[11] investigated the relationship between data-factor marketization and industrial land-use efficiency. Collectively, these studies suggest that the economic effects of data factors may extend beyond individual firms and influence regional resource allocation and sustainable development. At the same time, regional studies face methodological challenges. Differences in local institutions, infrastructure, industrial structure, and government policies may affect both data-factor marketization and economic outcomes, creating potential policy confounding. Moreover, the non-rival nature of data means that its effects may cross administrative boundaries, making spatial spillovers an important issue for future research.

The sixth stream focuses specifically on the institutionalization of data as a factor of production, particularly in the Chinese context. Zhang *et al.* ^[5], Wu ^[13], and Dong *et al.* ^[16] examined the development of data trading platforms, market construction, and factor-market institutions. This literature provides an important empirical and institutional setting for understanding how a relatively new production factor can be incorporated into an existing economic system. The Chinese experience is particularly valuable because it demonstrates that recognizing data as a factor of production involves not only technological development but also institutional innovation, market construction, property-rights arrangements, and policy coordination. However, the existing Chinese literature remains relatively separated from the broader international theoretical literature on data economics. Greater theoretical integration would help determine whether findings from the Chinese institutional context reflect general mechanisms of data-factor marketization or are strongly dependent on country-specific institutional conditions.

7.1. Integrated findings from the existing literature

Taken together, the six streams of literature reveal a clear progression from conceptualization to marketization, governance, and value creation. The economic literature explains why data differs from traditional production factors, particularly because of its non-rival characteristics, potential externalities, and ability to reduce information and transaction costs. The marketplace literature then addresses the institutional mechanisms needed to transform these characteristics into actual economic transactions. Governance and privacy research further demonstrates that data circulation cannot be separated from questions of rights, security, privacy, and institutional legitimacy.

At the enterprise level, the literature suggests that the productive value of data depends heavily on complementary capabilities. In other words, data is not necessarily productive simply because firms possess large quantities of it. Its economic contribution depends on whether firms can integrate data with digital technologies, organizational processes, human capital, and managerial decision-making. This observation provides an important connection between the data-factor literature and the broader literature on digital

transformation.

The regional literature further suggests that the effects of data-factor marketization may extend beyond individual enterprises. Data can facilitate knowledge diffusion, improve resource allocation, support innovation, and contribute to green development. However, these effects are likely to be heterogeneous across regions because digital infrastructure, human capital, industrial structures, and institutional environments differ substantially. Therefore, data-factor marketization should not be regarded as a uniform policy intervention with identical effects across locations.

Overall, the literature indicates that the economic value of data emerges through a multi-stage mechanism:

Data resources → institutional recognition → market circulation → complementary capabilities → productive application → economic and social value creation.

This mechanism also explains why simply increasing the quantity of available data may not be sufficient to generate economic growth. Effective data-factor development requires the simultaneous development of market institutions, governance mechanisms, digital infrastructure, human capabilities, and organizational capacity.

7.2. Research gaps

Despite the rapid expansion of the literature, several important research gaps remain. First, the measurement of data as a production factor remains challenging. Traditional production factors such as labor and physical capital can generally be measured through relatively established statistical systems, whereas the economic value of data depends on its quality, timeliness, specificity, accessibility, and actual utilization. As a result, existing studies may capture the institutional availability of data without accurately measuring the quantity or quality of data actually used by firms.

Second, there is still insufficient understanding of the long-term performance of data markets. Existing studies have provided important theoretical and institutional insights, but empirical evidence regarding market liquidity, pricing efficiency, transaction quality, and the sustainability of data marketplaces remains limited. Future studies should therefore move beyond the question of whether data markets exist and examine whether these markets can allocate data efficiently over time.

Third, the relationship between data marketization and governance requires further investigation. Data marketization may increase the circulation and economic utilization of data, but excessive circulation may also increase privacy and security risks. Future research should therefore investigate how different combinations of property rights, privacy protection, security mechanisms, and market regulations influence the efficiency of data markets.

Fourth, more attention should be given to the complementary conditions underlying enterprise-level effects. Current research provides evidence that data-factor marketization is associated with digital transformation and innovation, but the mechanisms through which these effects occur require further clarification. In particular, future research could examine the moderating roles of digital infrastructure, human capital, organizational capabilities, artificial intelligence, and management practices.

Fifth, the spatial dimension of data-factor marketization deserves greater attention. Because data is non-rival and can be transmitted across geographical boundaries, the economic effects of data may generate spatial spillovers that traditional regional production-factor models do not fully capture. Future research should therefore consider both direct and indirect regional effects and investigate whether data-factor development reduces or increases regional disparities.

Finally, the existing literature would benefit from greater integration between Chinese institutional research and international data economics. The Chinese experience provides a valuable setting for studying institutional innovation and the construction of data-factor markets, but future studies should examine the extent to which these institutional mechanisms can be generalized to other countries and economic systems.

In summary, the existing literature has established that data can function as an economically meaningful production factor, but the transformation from data resource to productive factor and ultimately to economic value remains an incompletely understood process. Future research should therefore adopt a more integrated perspective that simultaneously considers economic characteristics, institutional arrangements, market mechanisms, complementary capabilities, and economic outcomes. Such an approach would help develop a more comprehensive theoretical framework for understanding the role of data in contemporary production systems. The specific information is presented in **Table 2**.

Table 2. Synthesis of the literature

Literature stream	Core question	Representative evidence	Main contribution	Remaining limitation
Data economics	Why is data economically distinctive?	Non-rivalry, externalities, digital cost reduction ^[1-4]	Provides microeconomic foundations	Often abstracts from institutional implementation
Data marketplaces	How can data be exchanged?	Lifecycle, specificity, transaction costs ^[12]	Explains market intermediaries and institutional needs	Limited evidence on long-run market performance
Governance & privacy	How can circulation be made legitimate and safe?	Privacy, rights, security, governance ^[5,14,15]	Shows market design and governance are interdependent	Rights bundles and cross-jurisdiction comparisons remain underdeveloped
Enterprise effects	What changes inside firms?	Digital transformation, innovation quality ^[6,7]	Identifies complementary capabilities and channels	Measurement of actual data use remains difficult
Regional effects	What changes across places?	Innovation, green development, land efficiency ^[8-11]	Connects marketization with regional allocation	Potential policy confounding and spatial spillovers
Chinese institutional research	How is data institutionalized as a factor?	Trading platforms, market construction, factor policy ^[5,13,16]	Provides a large institutional laboratory	Need stronger integration with international theory

8. Critical gaps and theoretical tensions

8.1. Conceptual ambiguity

The first unresolved issue is conceptual. The terms data, data resource, data asset, data element, and data factor are often used without a stable hierarchy. This creates measurement problems because empirical studies may use a policy variable, a digitalization index, a data-trading-platform dummy, or a firm-level digital-technology measure as a proxy for the same broad concept. These measures capture different stages of the data-value chain. A platform dummy measures institutionalization; a digitalization index measures adoption; a data-asset measure captures organizational control or recognition; and raw data volume measures availability. They should not be treated as interchangeable.

8.2. Measurement and valuation

Data valuation remains one of the most difficult areas. Traditional assets are often valued through expected cash flows, market transactions, or replacement costs. Data creates difficulties for all three approaches because

transactions are heterogeneous, data can be reused, and value is often jointly produced with algorithms and organizational capabilities. A dataset may have no independent value before cleaning but very high value after integration. Conversely, a dataset may lose value rapidly as it becomes obsolete. Future research should therefore distinguish between acquisition cost, replacement cost, transaction price, expected economic value, and social value.

8.3. Causal identification

Recent empirical work increasingly uses data-trading-platform establishment as a quasi-natural experiment^[6–11]. This is an important methodological improvement, but it also raises a common concern: the establishment of a platform may not be random. Cities and regions that establish platforms earlier may already possess stronger digital infrastructure, fiscal capacity, human capital, market institutions, or policy attention. Staggered DID, event studies, placebo tests, and instrumental-variable strategies can reduce but not automatically eliminate these concerns. Future research should exploit more granular institutional shocks, transaction-level data, firm-level exposure measures, and cross-market variation.

8.4. Complementarity and the “data productivity paradox”

Data’s productive effect may be difficult to observe when complementary capabilities are weak. A firm may acquire a large dataset but lack analytical talent, computing infrastructure, organizational routines, or management systems. This provides a possible explanation for heterogeneous effects across regions and firms. The literature should therefore move from a one-factor model toward a complementarity framework in which data interacts with AI, human capital, capital investment, and organizational transformation.

8.5. Marketization versus governance

A further tension concerns the relationship between marketization and governance. Some discussions implicitly treat regulation as a constraint on data circulation, whereas others emphasize governance as a prerequisite for trust. The literature increasingly supports the second view. Without rights, standards, security, and enforcement, market transactions may be too risky or costly. The research challenge is therefore not to identify the ‘right amount’ of regulation in isolation, but to understand which governance arrangements reduce harmful externalities while preserving productive data use.

8.6. Distributional effects

Most studies focus on efficiency, productivity, and innovation, but distributional effects are less developed. Data-factor marketization can change the relative demand for skills, the bargaining position of workers, and the distribution of returns between data producers, platforms, intermediaries, and users. It may also reinforce regional advantages if high-capability regions can convert data into value more effectively than less-developed regions. Future research should therefore incorporate labor-market, income-distribution, and regional-inequality perspectives. The specific information is shown in **Figure 3**.

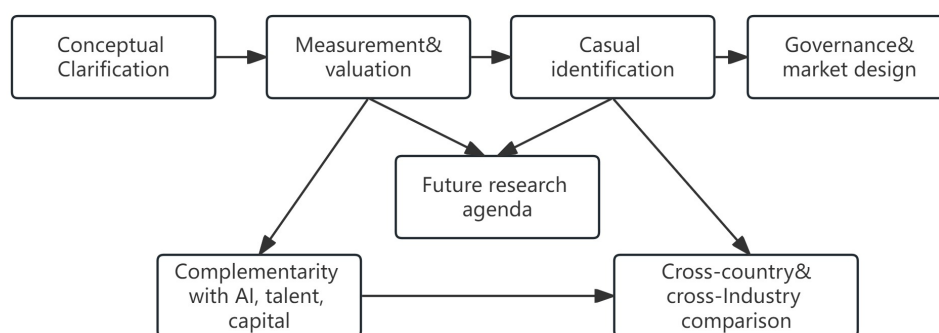


Figure 3. Research agenda for the next stage of data-factor research

9. Future research agenda

9.1. Build a unified conceptual architecture

Future research should establish a clearer hierarchy among data resource, data asset, data element, and data factor. One possible architecture is to treat data resource as the availability stage, data asset as the organizational and controllable stage, data element as the allocatable unit within a market or institutional system, and data factor as the broader economic role of data in production and allocation. Such a hierarchy would reduce conceptual overlap and improve measurement.

9.2. Develop multidimensional data-factor indicators

Rather than relying on a single proxy, researchers should develop multidimensional indicators covering supply, quality, access, market activity, infrastructure, governance, and productive use. For example, a regional index could separately measure data availability, data-service capacity, trading institutions, interoperability, security governance, and enterprise utilization. Firm-level research could combine data intensity with actual data-related investment, personnel, transactions, and organizational capabilities.

9.3. Advance valuation research

Data valuation should move toward scenario-based and use-case-based methods. Instead of asking “What is the price of this dataset?” research could ask how much incremental value the dataset creates for a particular application relative to a credible counterfactual. This approach is more consistent with the context dependence of data and could connect accounting valuation, managerial economics, and industrial organization.

9.4. Strengthen causal research

Future empirical work should combine institutional shocks with detailed measures of exposure. The establishment of a data exchange is a useful treatment, but actual exposure to data markets varies across firms and industries. Transaction-level data, data-service procurement, public-data access records, and firm-level data investment could provide stronger treatment measures. Spatial spillovers should also be modeled because data is highly transferable across geographic boundaries.

9.5. Examine AI–data complementarity

The rise of generative AI makes the data-factor question even more important. AI systems depend on data for training, fine-tuning, evaluation, retrieval, and application. At the same time, AI changes the marginal value of data by improving the ability to extract patterns from large datasets. Future research should therefore examine whether AI acts as a complement that raises the return to data, a substitute that reduces the need for certain data inputs, or a technology that changes the distribution of data value across firms.

9.6. Compare institutional models

China’s institutional construction provides a valuable case, but the theoretical implications should not be limited to one national context. Comparative research could examine China alongside the European Union’s data-governance framework, U.S. platform-oriented data markets, and other emerging institutional models. The key comparative question is which combinations of rights, standards, platforms, privacy rules, and competition policies generate efficient and inclusive data circulation.

9.7. Expand the sustainability and social-welfare dimension

Recent studies already link data-factor marketization to green innovation, urban environmental outcomes, and resource-use efficiency^[9–11]. Future reviews and empirical studies should integrate these outcomes with social welfare, inclusion, and distribution. The ultimate value of data factors should not be assessed only by market transaction volume or enterprise productivity; it should also consider whether data institutions improve the efficiency, inclusiveness, resilience, and sustainability of economic systems.

9.8. Implications for policy and management

The literature also has practical implications. For policymakers, the central task is not merely to increase the number of data exchanges or the volume of data transactions. Market infrastructure should be accompanied by standards, rights clarification, security mechanisms, interoperability, competition safeguards, and mechanisms for resolving disputes. Public policy should also pay attention to complementary conditions, including digital infrastructure, data literacy, human capital, and support for small and medium-sized enterprises. For managers, the key lesson is that purchasing or accumulating data does not guarantee value creation. Firms need data governance capabilities, analytical talent, appropriate organizational routines, and clear business use cases. Managers should evaluate data investments according to incremental economic value rather than dataset volume alone. For researchers, these practical issues reinforce the need to distinguish data availability from data utilization and data utilization from data-generated value.

9.9. Toward an integrated theory of data-factor markets

Taken together, the literature suggests that no single disciplinary perspective is sufficient. Economics provides theories of non-rivalry, externalities, information asymmetry, and market power. Institutional economics explains why rights and transaction rules matter. Management research highlights complementary capabilities and organizational change. Information systems research explains how data infrastructures and analytics create operational value. Law and public policy clarify privacy, security, and regulatory constraints. Accounting research contributes to measurement, recognition, valuation, and disclosure. An integrated theory should therefore treat data-factor markets as socio-technical institutions rather than pure markets. Such a theory would explain how data characteristics interact with institutions, platforms, firms, and technologies to determine who

can access data, how data is priced, how value is created, and how returns are distributed.

9.10. Toward a data-factor production-function perspective

The next generation of research may benefit from integrating data into production-function analysis while avoiding the assumption that data is simply another additive input. A useful starting point is a complementarity formulation in which output depends on physical capital, labor, conventional intangible capital, data, and the interaction terms among data, AI, organizational capability, and human capital. Such a framework would allow researchers to ask whether the marginal productivity of data rises with digital capability, whether data and AI are strategic complements, and whether organizational redesign is necessary for data to generate measurable returns. It would also help explain why empirical estimates differ across industries and regions. The challenge is empirical measurement: researchers need credible measures of data use rather than relying exclusively on platform establishment or broad digital-economy indices. Firm-level surveys, administrative records, data procurement contracts, data-related investment, and information-system logs could become valuable sources for future production-function research.

9.11. Data factors and spatial inequality

The regional literature also suggests that data-factor marketization may have asymmetric effects. Regions with stronger digital infrastructure, higher R&D intensity, deeper capital markets, and better human capital can convert data access into innovation more effectively than regions with weaker complementary conditions. Ouyang and Hu ^[8] found stronger innovation effects in regions with higher economic development, marketization, R&D investment, and baseline innovation capacity. Similar heterogeneity appears in enterprise-level studies. Zhan *et al.* ^[6] reported stronger digital-transformation effects in major cities and regions with stronger intellectual-property protection, while Xu *et al.* ^[7] found stronger innovation-quality effects in large and eastern cities. These findings raise a classic cumulative-advantage question: if data markets primarily benefit places that already possess complementary capabilities, marketization could initially widen regional gaps even while increasing aggregate efficiency. A complete evaluation of data-factor policy should therefore distinguish aggregate gains from distributional consequences.

9.12. Data marketization and labor and organizational change

A further underdeveloped dimension is the labor-market and organizational consequence of data-factor marketization. Data-intensive production changes the relative importance of analytical skills, managerial coordination, and routine information processing. When data become easier to access and combine, firms may reorganize tasks around real-time monitoring, prediction, and algorithmic decision support. The resulting productivity gains may increase demand for highly skilled workers while reducing demand for some routine tasks. This possibility is consistent with the broader literature on digital transformation, but data-factor marketization adds an institutional dimension: the availability and price of external data can influence the pace and direction of organizational restructuring. Recent evidence on Chinese listed firms suggests that data-factor marketization can increase firms' labor income share through human-capital upgrading, resource-allocation efficiency, and financing-constraint alleviation. Such findings indicate that the labor consequences of data markets need not be uniformly negative. Instead, they may depend on whether firms use data to substitute for labor or to augment worker capabilities.

10. Conclusion

This review synthesizes a rapidly developing body of research on data as a new factor of production and identifies a transition from conceptual recognition toward institutionalization and empirical evaluation. The literature establishes several important points. First, data is economically distinctive because it is non-rival, highly reusable, relational, and deeply complementary with technology and organizational capabilities. Second, these characteristics make data markets fundamentally different from conventional factor markets: access rights, quality, privacy, security, interoperability, and externalities matter alongside price. Third, data-factor marketization should therefore be understood as an institutional process that makes data more allocatable and usable, rather than simply as the creation of trading platforms. Fourth, emerging empirical evidence suggests that marketization can affect digital transformation, innovation, green development, and resource allocation, but these effects are heterogeneous and dependent on complementary capabilities.

The review also reveals a shift in the research frontier. The central question is no longer whether data has economic value. Instead, researchers increasingly ask how institutions convert potential data value into realized economic value, who captures that value, and under what conditions data circulation creates social gains rather than market failures. This shift creates an opportunity for interdisciplinary research that combines digital economics, institutional economics, management, information systems, accounting, law, and public policy.

The integrated framework proposed in this paper places data characteristics at the foundation, institutional conditions as the enabling layer, market mechanisms as the allocation layer, complementary capabilities as the conversion mechanism, and productivity, innovation, digital transformation, and sustainability as major outcomes. This framework can help reconcile international research on the economics of data with China-focused research on data-factor marketization. Future work should focus particularly on measurement, valuation, causal identification, AI–data complementarity, distributional consequences, and comparative institutional analysis. Such research will be essential for understanding whether data can become not only a new factor of production, but also a foundation for more efficient, innovative, and sustainable economic systems.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Goldfarb A, Tucker C, 2019, Digital Economics. *J. Econ. Lit.* 57, 3–43. <https://doi.org/10.1257/jel.20171452>.
- [2] Jones CI, Tonetti C, 2020, Nonrivalry and the Economics of Data. *Am. Econ. Rev.* 110, 2819–2858. <https://doi.org/10.1257/aer.20191330>.
- [3] Bergemann D, Bonatti A, Gan T, 2022, The Economics of Social Data. *RAND J. Econ.* 53, 263–296. <https://doi.org/10.1111/1756-2171.12407>.
- [4] Acemoglu D, Makhdoumi A, Malekian A, et al., 2022, Too Much Data: Prices and Inefficiencies in Data Markets. *Am. Econ. J. Microecon.* 14, 218–256. <https://doi.org/10.1257/mic.20200200>.
- [5] Zhang X, Shi Y, Sun K, et al., 2026, A Bibliometric Review of Data Factor Market: Contrasting Chinese and International Literature from a Governance Perspective. *J. Sci. Technol. Policy Manag.* 17, 1184–1208. <https://doi.org/10.1108/JSTPM-10-2023-0169>.
- [6] Zhan Y, Huang J, Yi Y, et al., Data Elements Marketization and Enterprise Digital Transformation. *Econ. Anal. Policy*

- 88, 992–1007. <https://doi.org/10.1016/j.eap.2025.10.011>.
- [7] Xu N, Xu L, Yan XW, 2025, Data Factor Marketization Empowering Enterprise Innovation Quality: New Evidence from Chinese Patent Citations. *Int. Rev. Econ. Financ.* 103, 104433. <https://doi.org/10.1016/j.iref.2025.104433>.
 - [8] Ouyang Y, Hu M, 2026, Data Factor Marketization and Regional Innovation: Evidence from the Establishment of Data Trading Platforms in China. *Technol. Soc.* 87, 103371. <https://doi.org/10.1016/j.techsoc.2026.103371>.
 - [9] Song Y, Xu C, 2025, Marketization of Data Elements and Corporate Green Innovation: Evidence from the Establishment of Data Trading Platforms in China. *Sustainability* 17, 10980. <https://doi.org/10.3390/su172410980>.
 - [10] Li H, Du X, Yan XW, et al., 2024, Digital Transformation and Urban Green Development: Evidence from China's Data Factor Marketization. *Sustainability* 16, 4511. <https://doi.org/10.3390/su16114511>.
 - [11] Chen W, Li S, 2025, Data Factor Marketization and Urban Industrial Land Use Efficiency: Evidence from the Establishment of Data Trading Platforms in China. *Sustainability* 17, 2753. <https://doi.org/10.3390/su17062753>.
 - [12] Huang L, Dou Y, Liu Y, et al., 2021, Toward a Research Framework to Conceptualize Data as a Factor of Production: The Data Marketplace Perspective. *Fundamental Research* 1, 586–594. <https://doi.org/10.1016/j.fmre.2021.08.006>.
 - [13] Wu J, 2026, From Fragmentation to Integration: The Structural Transformation and Maturation Mechanism of Data Factor Markets in China. *Economies* 14, 252. <https://doi.org/10.3390/economies14070252>.
 - [14] Acquisti A, Taylor C, Wagman L, 2016, The Economics of Privacy. *J. Econ. Lit.* 54, 442–492. <https://doi.org/10.1257/jel.54.2.442>.
 - [15] Goldfarb A, Tucker CE, Eds. 2024, *The Economics of Privacy*; University of Chicago Press: Chicago, IL, USA.
 - [16] Dong L, Zhang Y, Liu X, 2025, From Factor Cognition to Framework Construction: A Review of the Research on Data Production Factors from the Perspective of New Quality Productive Force. *Sci. Res. Manag.* 46, 13–24. <https://doi.org/10.19571/j.cnki.1000-2995.2025.09.002>.
 - [17] Li J, Gao L, Shi J, 2025, Too Good to Be True? China's New Conceptual Scheme for Data Property Rights. *Telecommun. Policy*. <https://doi.org/10.1016/j.telpol.2025.103145>.
 - [18] Yan H, Zhao Z, Zheng Y, et al., 2025, Evaluating the Energy-Saving and Efficiency-Enhancing Potential of Data Factor Marketization: Empirical Evidence from 270 Cities in China. *J. Environ. Manag.* 378, 124686. <https://doi.org/10.1016/j.jenvman.2025.124686>.
 - [19] Lu Y, Wang YB, 2024, Data Elements Marketization and the Development of New Quality Productive Forces: Causal Inference Based on Double Machine Learning. *Urban Probl.* 348, 80–90. <https://doi.org/10.13239/j.bjsshkxy.cswt.240707>.
 - [20] Brynjolfsson E, McElheran K, 2016, Data in Action: Data-Driven Decision Making in U.S. Manufacturing. U.S. Census Bureau Center for Economic Studies Working Paper CES-WP-16-06. <https://doi.org/10.2139/ssrn.2722502>.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.