

# The Impact of New Quality Productive Forces on Data Application Methods in Social Sciences: Postprint

**Authors:** Liu Shiguo

**Date:** 2025-03-20T00:00:00+00:00

## Abstract

In the positivist research paradigm, data is crucial for verifying theories and discovering patterns. Social science research has gradually increased its emphasis on data, and data application methods have progressively developed. Throughout this process, productive forces have played a vital role. Before the mid-19th century, social science research primarily relied on traditional speculative and deductive methods, with limited data application. Around the 1950s, the behavioralist revolution he...

## Full Text

### The Impact of New Quality Productive Forces on Data Application Methods in Social Sciences

**Liu Shiguo** (Institute of World Economics and Politics, Chinese Academy of Social Sciences, Beijing 100732)

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2025.01.04

In the positivist research paradigm, data is crucial for verifying theories and discovering patterns. Social science research has gradually increased its emphasis on data, and data application methods have progressively developed. Throughout this process, productive forces have played a vital role. Before the mid-19th century, social science research primarily relied on traditional speculative and deductive methods, with limited data application. Around the 1950s, the behavioralist revolution heightened social science's attention to data, and disciplines such as political science, sociology, and education began emphasizing data collection and organization, using statistics, formal models, and computer simulations to conduct research, thus initiating the rise of quantitative analysis. In the latter half of the 20th century, the development of computers and information technology propelled researchers to begin using massive amounts of

data for applied research, promoting the widespread application of quantitative social science methods.

Currently, new quality productive forces embody the major trends and currents of the new round of technological revolution and industrial transformation in today's world, and are profoundly influencing and will continue to influence the development of data application methods in social sciences.

First, they have greatly expanded the sources, types, and forms of social science data, pushing social science research from traditional small-sample, static, structured data toward multi-source, dynamic, multimodal data. Data sources have expanded from "limited channels" to "full-domain coverage." Traditional data mainly came from surveys, official statistics, and documentary archives. New quality productive forces have made data sources more diversified and real-time through technological means, bringing social media and online behavior data, IoT and sensor data, online platforms and digital footprints, blockchain and distributed ledgers, among others. These forces have pushed social science data types from single text and numerical data to multimodal, unstructured data that more closely approximates real social scenarios. Traditional data was primarily structured, such as statistical tables, questionnaire scales, and archival records. New technologies have given rise to unstructured text data, image and video data, spatiotemporal trajectory data, biological and physiological data, and virtual space data. They have also pushed data forms from static "snapshots" to dynamic, real-time, interactive "stream data." Through 5G and edge computing, real-time acquisition of social media public opinion fluctuations and financial market sentiment changes is possible. For example, user conversations with AI chatbots become human-machine collaborative generative data, and generative AI simulates social experiments to produce synthetic and simulation data.

Second, they have improved full lifecycle data management capabilities, laying a solid data foundation for social science data application methods. Perception and connection technologies have enhanced data acquisition capabilities. IoT hardware devices such as sensors and smart terminals collect physical world data in real time. 5G/6G communication technologies support massive device interconnection, enabling high-density data real-time transmission. Satellite remote sensing and drones provide global geographic spatial data acquisition capabilities. Edge computing technology performs preliminary processing at the data source, reducing data transmission pressure. AI-driven intelligent perception technologies such as computer vision and speech recognition automatically extract structured data. Distributed and automated technologies provide powerful data collection capabilities. Elastic resource pools such as cloud computing platforms support large-scale data convergence. Edge nodes with distributed architecture enable localized data caching and preprocessing. Stream data processing technologies collect and process dynamic data streams in real time. Automated crawlers and APIs automatically scrape publicly available network data. Intelligent processing and standardization enhance data organization

capabilities. High-performance computing clusters rapidly clean, annotate, and transform heterogeneous data. AI chips accelerate data preprocessing, such as image denoising and text segmentation. Natural language processing technologies automatically classify, deduplicate, and perform entity recognition. Data lakes and data middle platforms unify storage formats, enabling cross-source data fusion. Data storage technologies ensure high capacity and high availability of data. Distributed storage systems support low-cost petabyte-level data storage. New storage media such as quantum storage break through traditional capacity limitations. Blockchain technology ensures data immutability. Through AI algorithms, tiered management of hot and cold data is implemented, thereby optimizing storage resource allocation. Data dissemination enables secure sharing and efficient distribution of data. Content delivery networks with global nodes improve data access speed. Quantum communication networks guarantee absolute security of data transmission. Federated learning and privacy computing technologies enable data to be “usable but invisible,” such as in cross-institutional sharing of medical data. Smart contracts and data marketplace technologies enable automated data transactions through blockchain platforms.

Third, they have profoundly reshaped the methodology of social science data application, pushing social science research from “small-sample induction” to “full-sample computation,” from “static analysis” to “dynamic deduction,” from “manual interpretation” to “intelligent decision-making,” and even triggering a paradigm shift from “theory-driven” to “data-driven.” These forces have pushed the research process from linear to closed-loop iteration. The traditional process is generally “hypothesis—data collection—analysis—verification.” New quality productive forces drive innovation in research processes. “Real-time data feedback loop”: IoT sensors and edge computing enable real-time linkage of “data collection—analysis—strategy adjustment.” AI-assisted hypothesis generation: using natural language processing to mine literature databases to generate research hypotheses. Simulation verification front-loading: constructing virtual social systems through digital twin technology to preview policy effects. For example, the United Nations Development Programme uses “satellite imagery + AI algorithms” to monitor facility changes in poverty-stricken areas in real time, dynamically adjusting aid strategies. They have pushed analytical tools from statistical software to intelligent engine upgrades. Traditional analytical tools relied on SPSS, Stata, and other statistical software to process structured data, making it difficult to cope with unstructured, high-dimensional data, while new technologies enable the following empowerment. Multimodal data fusion analysis: knowledge graphs integrate character relationships, event spatiotemporal trajectories, and text semantics, such as visualizing interest networks in anti-corruption research. Complexity science tools: applying institution-based modeling to simulate the emergence of group behavior, such as multi-agent simulation of rumor spread during epidemics. Causal inference revolution: using reinforcement learning to construct counterfactual reasoning models, such as evaluating the long-term causal effects of educational poverty alleviation poli-

cies. They have realized innovation from laboratory to “social laboratory.” The bottleneck of traditional experimental methods lies in high costs of field experiments and difficulty controlling variables, with laboratory environments detached from real social scenarios, while new quality productive forces have brought full-domain social experimental fields, mixed-reality experiments, and synthetic data experiments. For example, Stanford University replicated real cities in the virtual world “Second Life” to study collective action logic in disaster emergency response. They have pushed collaboration models from closed to open transformation. Traditional scientific research collaboration has barriers such as data silos and opaque methodologies, while new quality productive forces have given rise to new collaborative ecosystems. Blockchain provenance and traceability: ensuring research processes are immutable and open-source community co-construction, such as building social science model libraries to share pre-trained models. Crowdsourced data annotation: such as Amazon Mechanical Turk combined with AI automatic quality inspection, completing multi-language social survey annotation at low cost. One successful case is the Global Climate Action Tracking project, which integrates satellite data, corporate carbon emission reports, and NGO investigations, achieving cross-national joint analysis through blockchain rights confirmation. Decision support has leaped from empirical judgment to data intelligence. Traditional decision-making relies on expert experience and lagging statistical indicators, while intelligent decision systems are based on wide-area information systems and AI models. For example, social risk early warning systems apply natural language processing and time series analysis to predict social risks, and policy simulation deduction engines evaluate the impact of minimum wage adjustments on industrial chains based on multi-agent modeling.

It should be noted that the impact of new quality productive forces on social science data application methods is currently in its early stages, and more profound and extensive transformations are yet to come.

[Author Bio] Liu Shiguo, male, researcher, research direction: world economic statistics and open macro econometrics, Email: liusg@cass.org.cn.

---

## Exploring from Practice, Improving the Mechanism for Data Asset Price Formation

**Wang Jian Dong** (National Development and Reform Commission Price Monitoring Center, Beijing 100837)

**Sun Jing** (Peking University School of Engineering, Beijing 100871)

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2025.01.05

In the digital economy era, data has become a key production factor. Establishing and improving the data asset pricing mechanism can provide strong support for promoting data factor circulation, revitalizing data assets, and unleashing data value. Currently, the phenomenon of “thousand people, thousand prices”

for data assets is relatively common. It is still necessary to support market practice, strengthen policy guidance, promote the construction of the data factor market system, and improve the market price formation mechanism for data assets.

## 1 The Data Asset Price Formation Mechanism Is Still in the Exploration Phase

The data factor is a major theoretical innovation first proposed by China, and data asset pricing is one of the global challenges without mature experience worldwide, requiring continuous exploration.

### 1.1 Data Asset Pricing Provides Support for Activating Data Factor Potential

Currently, all walks of life have accumulated large amounts of “sleeping” data assets. With the advancement of practices such as data resource capitalization, the work of evaluating and pricing data assets owned by enterprises and confirming them as assets has been put on the agenda. New quality productive forces are centered on the deep application of data factors, optimizing production processes and innovating business models through digital technology, and the market pricing of data assets is precisely the key prerequisite for unleashing this potential. Exploring market pricing of data assets can, on the one hand, prompt data supply entities to form a deeper understanding of data asset value, fully mobilize their enthusiasm and initiative, and strengthen data asset protection. In this way, “sleeping” data assets will be revitalized, helping achieve the goal of maintaining and increasing data asset value and improving the quality of data supply. On the other hand, scientific and reasonable market pricing of data assets will also promote compliant circulation of data factors in the market, enhance the professional operation capabilities of data merchants and third-party professional service institutions, help build standardized and efficient data trading venues, and optimize data resource allocation.

### 1.2 Basic Principles for Data Asset Market Pricing Are Gradually Being Established

Since the Fourth Plenary Session of the 19th CPC Central Committee officially listed data as a production factor, subsequent documents such as “On Building a More Perfect Market-oriented Allocation System for Factors”...

[Author Bio] Wang Jian Dong, male, researcher, research direction: data pricing, big data analysis, data factor governance, Email: wangjd\_{wyzx}@126.com; Sun Jing, female, assistant librarian, research direction: data analysis, scientific evaluation, etc., Email: sunjingcoe@pku.edu.cn.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*