ISSN 2616-5902 Vol. 6, Issue 8: 95-100, DOI: 10.25236/AJBM.2024.060815

Reflections on the Integration of Government Statistics and the Development of Big Data

Peng Ruiqin

Linyi Statistical Survey Center, Linyi, Shandong Province, China
pengruiqin@163.com

Abstract: *In the era of rapid development of modern information technologies such as the Internet of Things, big data, cloud computing, and AI, it is a crucial topic for government statistics to consider and actively address how to effectively utilize big data, a valuable social resource, and leverage big data thinking to drive government statistical reform. Against the backdrop of the "Big Data Era," this article explores the theoretical and practical aspects of why and how government statistics should integrate with big data, focusing on the inherent logic, external situation, basic principles and fundamental ideas of integrating government statistics with big data development.*

Keywords: *Government Statistics, Big Data, Integration, Reflection*

1. Introduction

Statistics is an essential foundational task for economic and social development, providing critical data for national macro-control and scientific decision-making. In today's era of rapid advancements in modern information technologies such as the Internet of Things, big data, artificial intelligence, and cloud computing, government statistics must contemplate how to accurately perceive the challenges big data presents, how to effectively utilize big data as a valuable social resource, and how to employ big data thinking to advance government statistics reform. This will enable better fulfillment of the three main functions of government statistical information, consultation, and supervision.

2. The Intrinsic Logic of Integrating Big Data Development with Government Statistics

The fundamental task of statistics is to conduct statistical surveys and analyses on the socio-economic development, provide statistical data and consulting opinions, and enforce statistical supervision. The prerequisite for accomplishing this fundamental task is to obtain statistical information, primarily statistical data, through statistical surveys. Integrating big data development with government statistics means proactively connecting with big data using an open and inclusive mindset and a concept of integration and innovation. This involves borrowing the thinking, organization, and technology used in the collection, integration, development, and application of big data to innovate statistical concepts, systems, mechanisms, methods, and processes, thereby fully realizing and enhancing the functions and roles of government statistics.

From the perspective of demand, international organizations have surveyed the demand for government statistical data. The results show that government departments are the largest users of statistical data, especially macroeconomic datas[1], followed by enterprises, research institutions, and the general public. For big data, enterprises in specific fields and industries, such as transportation, logistics, e-commerce, and finance, have a higher demand.

From the perspective of supply, the producers of government statistical data are legally authorized government statistical agencies, including comprehensive government statistical agencies at all levels (such as statistical bureaus and survey teams) and departmental statistical agencies. In contrast, the producers of big data encompass the entire society, including party and government agencies, enterprises, social organizations, as well as households and individuals.

From the perspective of production processes, the production of government statistical data follows a unified and standardized work plan, implemented step-by-step from top-down task deployment to bottom-up feedback. The entire process includes ten stages: demand determination, statistical design, approval and filing, task deployment, data collection, data processing, data evaluation, data publication,

ISSN 2616-5902 Vol. 6, Issue 8: 95-100, DOI: 10.25236/AJBM.2024.060815

statistical analysis, and comprehensive evaluation. Big data, however, is a byproduct that emerges simultaneously with the activities of supply and demand during transactional or non-transactional activities.

From the perspective of data sources, government statistical data mainly derive from various censuses, routine surveys, sampling surveys, and administrative records formed through daily management organized by the government. These source data are often "secondary data" that have been integrated and processed. In contrast, big data originates from the "digital traces" of transactions and non-transaction activities among various organizations, households, and individuals. These source data are often "raw data" generated at the time of the activity.

From the perspective of data content, government statistical data focus on population, economy, society, resources, ecological environment, and science and technology. Big data, however, is much broader, more complex, and detailed, encompassing both transactional and non-transactional content such as production and operation, clothing, food, housing, transportation, medical care, education, research, management services, and also including communications, public opinion, preferences, and more.

From the perspective of data forms, government statistical data are mostly structured data expressed in two-dimensional tables, whereas big data consists largely of unstructured text, images, audio, video, and other multidimensional data.

From the perspective of data quality, statistical data, governed by statistical laws, regulations, and standardized reporting systems and business processes, are generally of high quality. In contrast, big data is often mixed with good and bad elements, commonly plagued by issues like data loss, volatility, anomalies, and lower continuity, stability, and verifiability.

From the perspective of data timeliness, government statistics are retrospective, reflecting data after the fact and hence lacking in timeliness. Big data, however, is real-time. In the internet era, data is ubiquitous, pervasive, continuously generated, and utilized by people.

From the perspective of data disclosure, government statistical data is proactively and periodically released by relevant government departments, with the content mainly comprising macro-aggregate data, supplemented by structural data by industry and region. In accordance with the "Statistics Law," individual and sensitive data are not disclosed. Conversely, big data has multiple disclosure entities, random disclosure times, more specific and detailed disclosure content, and a greater focus on micro- and meso-level industries, fields, and enterprises.

In summary, both government statistical data and big data are forms of data with inherent logical connections. However, they each have distinct characteristics and strengths in terms of data sources, content, forms, quality, timeliness, and effectiveness. As a crucial foundational work for national management, government statistics should play a leading role in data management and application. This requires fully leveraging the advantages of big data, including its broad sources, wide scope, rich content, detailed classification, strong timeliness, and low cost, to expand the domain, increase the volume, enhance the quality, and add value to the use of government statistical data. This constitutes the intrinsic logic of integrating big data development with government statistics.

3. The External Context of Integrating Big Data Development with Government Statistics

Some people think that with the wide application and rapid development of modern information technologies such as the Internet, cloud computing, and space remote sensing in various fields of national economy and society, the era of big data will accelerate the advent of the artificial intelligence (AI) era, and future statistics may become "unmanned statistics." Others have analyzed the challenges faced by government statistics in the era of big data from three aspects: the coverage and content of the national data system, the suppliers of the national data system, and the demanders of national data. The conclusion is that government statistics will exist for a long time but must confront the impacts and challenges brought by big data through technical, organizational, and methodological reforms.

3.1. The New Mission of Integrating Big Data Development with Government Statistics

The era of big data has endowed the government with a new mission of integrating big data development with statistics. In the big data era, countries around the world have rapidly increased their dependence on data, shifting the focus of national competition from capital, land, population, and

ISSN 2616-5902 Vol. 6, Issue 8: 95-100, DOI: 10.25236/AJBM.2024.060815

resources to big data. Ian Ayres, an econometrician at Yale University, stated in "Super Crunchers: Why Thinking-By-Numbers is the New Way to Be Smart" that "those who control data control the world." Like land, data assets are expected to become key strategic resources for national development. In the big data era, no organization or individual can stay out of it. Particularly, government statistics, which rely on data, should take the lead and bear the responsibility of actively adapting to, grasping, and leading the big data era. By fully utilizing big data, a "free" resource of industrial society, government statistics can better fulfill their tasks and mission.

3.2. New Requirements Arising from the Modernization of National Governance

The modernization of national governance has created new requirements for the integration of government statistics with big data development. Big data represents not only a technological revolution and economic change but also a transformation in national governance. Viktor Mayer-Schönberger, a professor at Oxford University, stated in his book "Big Data: A Revolution That Will Transform How We Live, Work, and Think" that "big data is a source of new insights and new value, and a means to transform markets, organizations, and the relationship between governments and citizens." In promoting the modernization of the national governance system and capacity, big data has shifted the basis of government decision-making from limited "sample data" to vast "all data." As government governance transitions from traditional authoritative governance to data governance, government statistics should adapt to the requirements of national governance modernization by continuously improving and enhancing themselves in areas such as "open statistics," "real-time statistics," and "total statistics."

3.3. New Solutions to the Current Challenges Faced by Government Statistics

The current challenges faced by government statistics have given birth to new solutions through the integration of big data development. Presently, government statistics are at an opportune period for transformation and development, yet there are contradictions at the grassroots level such as limited personnel, high mobility, and heavy workloads. Problems like "information islands," "data silos," multiple surveys, repeated surveys, and inefficient surveys between departments and specialties are frequent and difficult to resolve. However, big data has demonstrated effective solutions by expanding scope, field, and object coverage, refining content, providing quality feedback, optimizing processes, reducing costs, and improving timeliness, thus offering a new path for government statistics to overcome their current dilemmas.

3.4. New Support from Software and Hardware Development

The construction of software and hardware provides new support for the development of government statistics and big data. On the one hand, China ranks among the top in the world in terms of the number of infrastructures such as supercomputers and 5G networks. On the other hand, China’s large population and huge economic scale continuously enrich the "big database" with the constant flow of people, logistics, capital, and information[2]. Following the establishment of the national big data development strategy, comprehensive promotion is underway in areas such as development planning, platform construction, and industrial docking. Initiatives like "humans working while the cloud is computing," "information running more, people running less," and "getting things done in one go" have seen vivid practice in fields such as industrial design, public services, and administrative approval. The National Bureau of Statistics has signed strategic cooperation agreements on big data with enterprises such as Tencent, Alibaba, and Baidu, enabling direct access to massive data from e-commerce transactions, industry platforms, search engines, and other sources. Singapore’s Statistics Bureau’s practice of obtaining income tax and consumption tax databases through authorization also offers valuable lessons. The rapid improvement of software and hardware construction levels provides new support for the integration of government statistics and big data development.

4. The Basic Idea of Integrating Government Statistics with the Development of Big Data

The integration of government statistics with the development of big data should aim to construct a modern statistical investigation system, enhance government statistical efficiency, and focus on improving government statistical capabilities. This involves adhering to problem-oriented and innovative concepts, systematic planning and top-level design, government leadership and open sharing,

ISSN 2616-5902 Vol. 6, Issue 8: 95-100, DOI: 10.25236/AJBM.2024.060815

and key breakthroughs with steady progress. The goal is to accelerate the formation of an innovative development path where government statistics and big data merge, synergize, and add value.

4.1. Adhere to Problem Orientation and Concept Innovation

The value of data lies in its application, applicable to both statistical data and big data. Facing new tasks, requirements, the current dilemmas of government statistics, and the successful application of big data in statistics, it is essential for government statistics to re-examine traditional statistical resources, theories, and concepts.

First, establishing the value concept of big data as a strategic resource. Government statistics should regard big data as a fundamental strategic resource, an inexhaustible and increasingly valuable green resource, and actively incorporate it into the management and application of government statistical data.

Second, actively promoting theoretical innovation. The overall concept of big data is not equivalent to the overall statistical target, nor is it a random sample of the target population. The representativeness of the potential population in big data and the processes and methods of statistical inference using big data are incompatible with traditional statistical theories and methods, necessitating innovation in statistical theory and methodology. For example, in the field of price statistics, scanning price data and web-scraped price data have become new data sources for compiling price indices in many developed countries. In the face of the vast quantity and rapid updates of full-scale data for pricing products, it is necessary to break through the traditional theoretical framework of price index, which “compares the prices of homogeneous comparable products and services in a fixed-quantity basket across different periods”[3].

Third, accelerating the transition from "+Internet" to "internet plus" and then to big data thinking. From a statistical perspective, "+Internet" data refers to data that various sectors of society make available online, either actively or passively, for use by government statistical agencies. "Internet plus" data refers to online data that government statistical agencies obtain, store, analyze, and process without technical barriers. Big data from a statistical perspective encompasses data, technology, and thinking as an integrated whole[4]. The transition from "+Internet" to "internet plus" and then to big data represents a shift in the statistical functions of government statistics from "production-management" to "management-production." It is also a process of laying the foundation, building the infrastructure, and preparing the raw materials for statistical artificial intelligence.

4.2. Adhere to System Planning and Top-Level Design

Promoting the integration of government statistics and big data is a systematic project that requires system planning and top-level design.

First, develop a comprehensive plan for integrating government statistics with big data. This includes standardizing and integrating statistical operations, building a statistical cloud big data application platform, enhancing big data application innovation, and advancing big data statistical research. The plan should be based on current needs while anticipating future developments, and it should be formulated at the national level.

Second, pursue top-level design and innovative layout. The significant progress in Guiyang's big data industry in recent years can be attributed to its strategic top-level design. This design encompasses the entire lifecycle of big data industry development, from the front end to the back end, implementing a "combination punch" of organizational formation, institutional creation, and park establishment. For instance, Guiyang enacted the first local regulation on government data sharing and openness in China, the Guiyang Government Data Sharing and Openness Regulations, followed by the Guiyang Big Data Security Management Regulations and the Guiyang Health and Medical Big Data Application Development Regulations. These regulations ensure the standardized development of the big data industry through legal guarantees. Guiyang's experience in top-level design and innovative layout offers valuable insights for the integration of government statistics and big data development.

Third, strengthen organizational leadership and coordinate actions. The application of big data in the field of statistics is not limited to specific regions but is a nationwide unified action and a government initiative. It requires strengthened organizational leadership and pilot projects at the national level to integrate big data development into local government statistics.

ISSN 2616-5902 Vol. 6, Issue 8: 95-100, DOI: 10.25236/AJBM.2024.060815

4.3. Adhering to Government Leadership and Open Sharing

Given the diverse sources of big data, which span various sectors, industries, levels, and aspects of the economy and society, there are significant differences in data content, form, quality, period, application, and management. The development and utilization of such data involve multidisciplinary knowledge and skills, including computer science, applied mathematics, statistics, and economics. Consequently, relying solely on government statistical agencies is insufficient; there is a need to adhere to government leadership and promote open sharing.

First, improve the legal framework to ensure the security of big data sharing, as well as the protection of trade secrets, commercial interests, and personal privacy.

Second, establish a mechanism for big data openness and sharing at the institutional level. This can be achieved by fulfilling legal obligations, forming strategic partnerships, and promoting collaborative sharing between government statistical agencies and other big data producers.

Third, increase government investment. Integrating big data into government statistics requires the construction of appropriate hardware and platforms for data collection, storage, processing, and analysis. It also necessitates innovation in related theories, disciplines, technologies, and talent. The government should introduce policies and plans to encourage accelerated integration and sustain increased financial investment.

Fourth, leverage both domestic and international resources for application. The application of big data in statistics differs from traditional statistical data development models and exceeds the processing capabilities of existing statistical infrastructures. It requires specialized knowledge in model differentiation, machine learning algorithms, and model estimation, as well as expertise in natural language processing, audio signal processing, and image processing. Therefore, government departments, especially statistical agencies, should not only enhance their internal capabilities but also actively seek external assistance. By combining internal and external efforts, they can collectively enhance the capacity and level of big data application in government statistics in China.

4.4. Key Breakthroughs and Steady Progress

In the process of integrating big data, it is essential to prevent "big data arrogance," which can compromise the scientific rigor and meticulousness of government statistical work, thereby affecting its authority and public trust. Similarly, it is important to avoid hesitation and inaction that can lead to stagnation. Therefore, a balanced approach that emphasizes strategic planning, feasible execution, key breakthroughs, and steady progress is necessary.

By consolidating resources, concentrating efforts, and fostering collaboration, the integration can be effectively promoted.

This will create a cohesive development system within the government statistical system, ensuring smooth vertical and horizontal integration.

This will reinforce the foundation, clarify the objectives, and expedite the integration process through self-imposed pressure and accelerated implementation.

5. Conclusion

In short, by analyzing the internal logic and external conditions, the integration of big data and statistics is a requirement for the reform and development of statistics and an inevitable development of government statistics under the conditions of high-speed development of information technology. From the primitive statistics in the early electronic calculator period, to the networked direct reporting statistics in the Internet period, to the massive statistics in the current big data era, information technology has always been the main driving force to promote the statistical work. High-speed development of information technology has given birth to the "big data era", government statistics should seize the opportunity to actively use big data to promote the innovation of statistical methods and statistical means, strengthen system integration, data fusion, and realize the in-depth fusion of big data and government statistics to promote the digital transformation of government statistics, and continuously promote the application of big data in the work of government statistics.

ISSN 2616-5902 Vol. 6, Issue 8: 95-100, DOI: 10.25236/AJBM.2024.060815

References

[1] United Nations, (2006) The Third Edition of Handbook for Statistical Organizations, China Statistics Press, 15-19.
[2] Yu F D,(2018) The application, bottleneck and fusion path of big data in government statistics 2018(11), 3-10.
[3] The International Monetary Fund, (2008) Handbook of the Consumer Price Index, Theory and Practice, 12-21.
[4] He Q,(2015). Discussion on the core ideas of big data in the view of government statistics, The world of research, 2015(2), 50-53.