

## NEW DATA SOURCES AND METHODS FOR OFFICIAL STATISTICS

*Siddiqui Amir,*  
Data Scientist, T Systems;  
Mumbai, India

### **Abstract:**

In the age of big data, the landscape of official statistics is experiencing a transformative evolution marked by the introduction of new data sources and inference methods. This paper explores the dynamic landscape of disseminating statistical data through new data sources. The study, beginning with a focus on the crucial role of official statistics in influencing policy decisions and shaping societal understanding, recognizes the limitations of conventional methodologies. The research acknowledges the constraints of traditional methodologies. The investigation is dedicated to exploring new data sources such as administrative records, sensor data, and big data, coupled with contemporary inference techniques like machine learning and Bayesian approaches. The abstract outlines the research's emphasis on discovering new data sources, presenting a concise overview of emerging data streams that include metrics from social media, satellite imagery, and datasets generated by the Internet of Things (IoT).

### **I. Introduction**

The evolving landscape of data sources and inference methods is marked by the rapid expansion of digital technologies, ushering in an era where vast amounts of data are generated every second and accessible by organizations. Traditional approaches to official statistics, while foundational, are facing challenges in terms of availability, scalability, and comprehensiveness. Given this background, it is important to incorporate new data sources and new information in the official statistics. It's important to take a closer look at new data sources and advanced inference methods to extract meaningful insights from the vast and diverse datasets now available.

### **II. Importance of Official Statistics**

Official statistics serve as the foundation for making decisions based on evidence, creating policies, and gaining insights into society. Typically produced and endorsed by government agencies, these statistics offer crucial insights into a nation's economic, social, and environmental aspects. In the realm of governance, policymakers heavily depend on official statistics to guide their decisions, allocate resources wisely, and assess the impact of policies over time. They empower citizens by offering a comprehensive view of national and local trends, enabling informed participation in public discourse.

### **III. Traditional Approaches to Official Statistics**

Over the years, a variety of traditional approaches, such as survey methodology, census enumeration, administrative data collection, and sampling techniques, among others, have been utilized to collect, analyze, and disseminate statistics, offering valuable insights into economic, social, and environmental aspects. However, these approaches come with their own set of limitations. Conventional methods frequently encounter challenges related to availability and scalability, especially when handling

large volumes of data. As societal challenges expand in both scope and complexity, traditional statistical methods may encounter difficulties in providing insights that are both timely and comprehensive.

#### **IV. Advancement in Data Sources**

Emerging sources such as administrative records, sensor data, social media, and big data have garnered significant attention. Administrative records, which include data collected by government agencies for non-statistical purposes, offer a rich reservoir of information. Examples of administrative records include data related to taxation, healthcare, education, social services, and law enforcement. Leveraging administrative records in official statistics can enhance the accuracy, efficiency, and timeliness of data collection, offering valuable insights into various aspects of public life while minimizing the burden on respondents. Studies suggest that leveraging administrative records can enhance the accuracy and coverage of official statistics, addressing some of the limitations of traditional survey-based approaches. [1] Unlike traditional survey methods, administrative records offer a continuous and cost-effective stream of data, eliminating the need for periodic surveys.

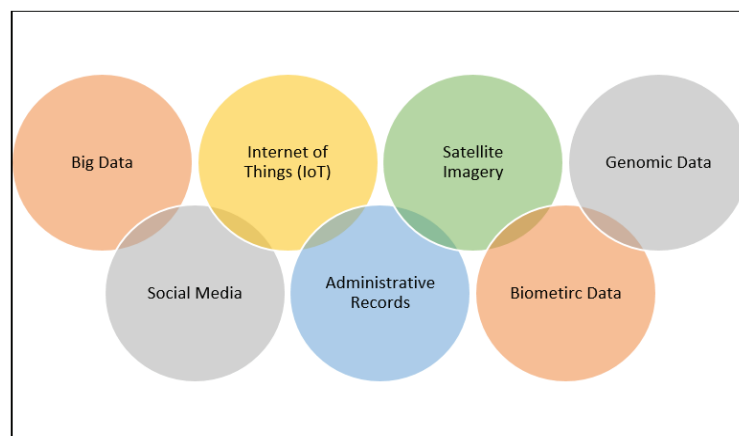
The advent of the Internet of Things (IoT) and sensor technologies has ushered in an era of real-time and granular data collection. In official statistics, leveraging sensor data and IoT offers the potential to capture dynamic societal changes in real time, providing a more nuanced understanding of trends and patterns.

Big data, characterized by vast and diverse datasets, offers an opportunity to integrate information from various sources. Data linkage techniques involve connecting datasets based on shared variables, facilitating the creation of comprehensive datasets for analysis.

Collaborative data-sharing initiatives among countries and organizations enable the pooling of resources and expertise. These collaborations foster the development of more comprehensive and globally relevant statistical datasets.

Remote sensing technologies and satellite imagery contribute geospatial data for official statistics. This includes monitoring land use changes, environmental conditions, and disaster impacts, offering a comprehensive understanding of geographic trends. [2] The inclusion of new

data sources can improve the relevance of official statistics by capturing emerging trends, insights from the combined information.



**Graphic showcasing a range of sources generating data**

## **V. Inference Methods**

In statistics, inference methods refer to techniques and approaches used to draw conclusions, make predictions, or gain insights from data. These methods are broadly categorized into:

1. Machine learning
2. Bayesian Inference

Machine learning, a subset of artificial intelligence, offers powerful tools for predictive modeling and imputation in the realm of official statistics. Predictive modeling involves utilizing historical data to forecast future trends or outcomes. Machine learning algorithms, ranging from decision trees to deep neural networks, excel at identifying patterns within vast datasets, providing more accurate predictions than traditional statistical models.

Bayesian inference, named after the Reverend Thomas Bayes, this approach uses probability theory to model uncertainty and quantify the strength of beliefs in various hypotheses. The core idea of Bayesian inference is to update the probability of a hypothesis based on new evidence.

Data fusion involves the integration of information from diverse sources to create a unified and comprehensive dataset. Data fusion methods include techniques for merging datasets, reconciling conflicting information, and deriving more accurate insights from the combined information.

## **VI. Challenges and Opportunities**

There are some challenges associated with integrating new data sources such as data quality and privacy concerns.

### **Challenges**

New data sources may vary in quality, leading to potential inaccuracies in official statistics. Implement rigorous data validation processes, establish quality standards, and collaborate with data providers to enhance accuracy. Increased reliance on diverse data sources raises privacy concerns, especially with sensitive information. Implement robust privacy protocols, anonymize data when possible, and adhere to legal and ethical frameworks to safeguard individuals' privacy. Integrating new data sources may face technological hurdles, including compatibility issues and data format disparities. Foster collaboration between data providers and statistical agencies, and develop standardized protocols for data sharing.

### **Opportunities**

New data sources, such as real-time sensor data and social media metrics, offer the potential for more timely statistical updates. Invest in technologies that enable faster data processing, explore automated data collection methods, and establish partnerships with entities providing real-time information. Diverse data sources provide opportunities for more granular insights into societal phenomena.

## **VII. Conclusion**

In the era of digital advancements, the changing role of official statistics reflects a dynamic interaction between established practices and modern innovation. The incorporation of emerging data sources and sophisticated inference methods establishes official statistics as a guiding source of understanding in the intricate

dynamics of contemporary societies. The significance of statisticians, policymakers, and researchers intensifies as they play a crucial role in navigating this shifting terrain. The digital age necessitates not only technical expertise but also a commitment to ethical principles, ensuring responsible data use with measures to safeguard privacy and promote transparency.

## References

1. Harnessing Administrative Data for a More Resilient Data and Statistical System, Asian Development Bank (ADB). URL: <https://www.adb.org/sites/default/files/publication/812946/part4-nsss.pdf>
2. UN Statistics Division. URL: [https://unstats.un.org/bigdata/task-teams/earth-observation/UNGWG\\_Satellite\\_Task\\_Team\\_Report\\_WhiteCover.pdf](https://unstats.un.org/bigdata/task-teams/earth-observation/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf)

## DERIVING NUMBERS FROM A CENTRAL BANK'S TEXTUAL RELEASES. THE CASE STUDY OF THE NATIONAL BANK OF UKRAINE TWITTER/X POSTS

*Szyszko Magdalena,*

Institute of Economics and Finance;  
WSB Merito University Poznan, Poland;

*Rutkowska Aleksandra,*

Department of Applied Mathematics;  
Poznan University of Economics and Business, Poznan, Poland;

*Motuzka Olena,*

Department of Management, Marketing and Public Administration;  
National Academy of Statistics,  
Accounting and Audit, Kyiv, Ukraine

For this study, we collected 552 Twitter/X posts published by the National Bank of Ukraine on its official profile in English. We collected all tweets published since June 6, 2019. The sample ends in August 2023. 488 posts were published before the invasion and 64 – after the Russian invasion.

The goal of the study is:

- to assess the sentiment of the NBU tweets in the comparative context (before the invasion and during the war – we assume less positiveness after the invasion);
- to assess the subjectivity of the NBU tweets in the comparative context (before the invasion and during the war – we assume less subjectivity after the invasion);
- assess the applicability of ChatGPT as a sentiment/subjectivity analysis tool compared with standard dictionary methods.