Popoola, Osuolale Peter; Adeboye, Olawale Nureni

Research Report

Fourth Industrial Revolution and Evolution of Data Science: Challenges for Official Statistics

Fourth Industrial Revolution and Evolution of Data Science: Challenges for Official Statistics

[1]Osuolale Peter POPOOLA and [2]Olawale Nureni ADEBOYE

1. Mathematics and Statistics Department, Adeseun Ogundoyin Polytechnic, Eruwa Oyo State, Nigeria
2. Department of Statistics, Osun State University, Osogbo, Osun State, Nigeria

Corresponding Author Email: osuolalepeter@yahoo.com

Abstract

Fourth Industrial Revolution is describes as exponential growth of several key technological fields' concepts, such as intelligent materials, cloud computing, cyber-physical systems, data exchange, the Internet of things and blockchain technology. At its core, data represents a post-industrial opportunity. The effects of technologies have provided new avenues of data for official statistics, which can then be harnessed through the power of data science. However, as data continue to grow in size and complexity; new algorithms need to be developed so as to learn from diverse data sources. The limitation of conventional statistics in managing and analyzing big data has inspired data analysts to venture into data science. Data Science is a combination of multiple disciplines that use statistics, data analysis, and machine learning to analyze data, and extract knowledge and insights from it. These swathes of new digital data are valuable for official statistics. This paper links industrial eras to the evolution of statistics and data; it examines the emergence of big data and data science, what it means, it benefits and challenges for official statistics

Keywords: *Industrial Eras, Data Evolution, Big Data Revolution, Data Science, Official Statistics*

1.0 Introduction

The fourth industrial revolution is a name for the current trend of automation and data exchange in manufacturing technologies, including cyber-physical systems, the Internet of things; cloud computing and cognitive computing creating the smart factory. It is a term coined by Klaus Schwab, founder and executive chairman of the World Economic Forum, describes a world where individuals move between digital domains and offline reality with the use of connected technology to enable and manage their lives (Miller 2015). It represents a new stage in the organization and control of the industrial value chain, data production and usage. The speed and measure of the changes coming about by the fourth industrial revolution are not to be ignored. These changes will bring about shifts in data production, data usage by official statisticians, shifts in wealth, and knowledge. Only in being knowledgeable about these changes and the speed in which this is occurring can official statisticians can advance in data production, data usage knowledge and technologies benefit all. Although each industrial revolution is often considered a

separate event, together they can be better understood as a series of events building upon innovations of the previous revolution and leading to more advanced forms of production. The first industrial revolution started in 1760 with the invention of the steam engine. The steam engine allowed the transition from farming and feudal society to the new manufacturing process. This transition included the use of coal as the main energy while trains were the main means of transportation. Textile and steel were the dominant industries in terms of employment, value of output, and capital invested. The second industrial revolution began in 1900 with the invention of the internal combustion engine. This led to an era of rapid industrialization using oil and electricity to power mass production. The third industrial revolution started in 1960 and was characterized with the implementation of electronics and information technology to automate production. Under the old ways, making things involved screwing or welding lots of parts together. The fourth industrial revolution now involves computer generated product design, three dimensional (3D) printing, big data revolution, data science, cloud computing, pattern recognitions, and image processing to extract knowledge and insights from it. These swathes of new digital data are as valuable for official statistics.

The possibilities of billions of people connected by mobile devices, with unprecedented processing power, storage capacity, and access to knowledge, are unlimited. And these possibilities will be multiplied by emerging technology breakthroughs in fields such as artificial intelligence, robotics, the Internet of Things, autonomous vehicles, 3-D printing, nanotechnology, biotechnology, materials science, energy storage, and quantum computing, and data science.

Already, artificial intelligence is all around us, from self-driving cars and drones to virtual assistants and software that translate or invest. Impressive progress has been made in AI in recent years, driven by exponential increases in computing power and by the availability of vast amounts of data, from software used to discover new drugs to algorithms used to predict our cultural interests. Digital fabrication technologies, meanwhile, are interacting with the biological world on a daily basis. Engineers, designers, and architects are combining computational design, additive manufacturing, materials engineering, and synthetic biology to pioneer a symbiosis between microorganisms, our bodies, the products we consume, and even the buildings we inhabit.

1.2 Statistics and Data Evolution

When the earth was formed several years ago, the early years were pretty chaotic. Eventually, oceans formed, and simple organisms began to evolve, followed by more complex plants, animals and, finally, man. Statistics have also been evolving into more complexes forms granted over a much shorter period of time. The earliest data collections took the form of census; sampling was eventually discovered and gave rise to surveys and then multi-topic surveys. Statistics is a scientific discipline that uses results from analysis of data to make decision in the face of uncertainty and with minimum risk. It deals with the production, collection, organization, and analysis of data, and interpretation and presentation of results. Statistical methods are used in almost all spheres of life and disciplines, especially in research. Statistics has made tremendous contribution to the humanity by providing data collection and production tools such as population census, sample survey based on random sampling method, and conducting controlled experiments to determine cause and its effect relationship, and randomized clinical Trial (RCT). It has changed the way industry used to control product quality by introducing quality control chart, total quality management (TQM) and sequential sampling method. Statistical models are used for analysis of variance, multiple regression analysis, time series analysis as well and forecasting and prediction. Statistical inference allows estimating unknown parameters, perform test of hypotheses etc. As the world changes, likewise, statistics has changed rapidly.

Data is a valuable asset to statistics, of a fact; one could easily say no data, no statistics. In 2006, Clive Humby (a British mathematician) coined the phrase “data is the new oil” about the availability of both resources: neither oil nor data is valuable in its raw state; rather, value is derived when it is gathered rapidly, completely, accurately and is connected to other relevant data. With real-time intelligence at their disposal, with data one can make informed decisions on the direction a nation is heading to. This is because data is equivalent to knowledge. Thus, owning good data serves as indisputable evidence or justification for a decision – it allows leaders to say, “we’ve done x because of y.” Essentially, good data is persuasive in and of itself. The other paths to follow include anecdotal evidence, assumptions, and abstract observation may potentially be easier but may consequently lead down a road of wasted resources. Gone are the days when assumptions and gut feelings steer the ship. Using data enables government of a nation to make less risky decisions based on facts provided by the data. Data production in statistics has equally changed from manual count of fingers, to counting using tallies, to a more

complexity of big data. At its core, data represents a post-industrial opportunity. Its uses have unprecedented complexity, velocity and global reach. Data is increasingly building up on who we are, who we know, where we are, where we have been and where we plan to go. As we move into the fourth industrial era, data will rule the world. Fourth industrial era is a digital technologies and data driven, internet dependent and satellite guided. Starting from driver-less vehicles to store-less shopping platforms and delivery of personalized services will be digitized based on data guided evidence. In fact, data revolution is already here, and we are increasingly being exposed to various technologies that are dependent on results from analysis and prediction of data. The role of statistics and computing algorithms in the process are crucial and it will continue to grow. The policymakers in government offices, official statisticians, health services, technology centers and business establishments are moving towards evidence-based decision-making which is predominantly guided by data synthesis and analytics.

2.0 Fourth Industrial Era and Data Revolution

Data is a set of values of subjects with respect to qualitative or quantitative variables. In its raw form, it is unorganized, but must be transformed, analyzed in order to be useful for policy making, monitoring and accountability and international comparison. It is an indispensable tool in statistics. Data revolution is already here, and it will continue to grow more and more at an accelerated rate in the days and months and years to come especially as the world move in to the fourth industrial era. Data are useful and potentially beneficial to mankind if the invaluable 'jewel' in the mess of data can be uncovered by using appropriate methods. Increasingly evolving statistical methods are the main vehicle to compile, explore and analyze raw and unformatted data and interpret the results leading to evidence-based decision-making.

The omnipresence of data in the daily lives of most people in the world gives rise and support to the view that data will change the world. With the unprecedented rate of data creation, and the increasing role data plays in most of our lives, it is easy to assume that the digital revolution could be the most important life-changing event of this era**.** Data revolution is already here, and it will continue to grow more and more at an accelerated rate in the days and months and years to come. In the contemporary world, there is no shortage of data. Rather we have a plethora of data almost everywhere. The amount of data the world produces every day is truly mind-boggling. There are 2.5 quintillion bytes of data created each day at our current pace, but that pace is only

accelerating with the growth of the Internet of Things. The Internet of Things refers to the billions of physical devices around the world that are now connected to the internet, all collecting and sharing data. Connecting all these different objects and adding sensors to them adds a level of digital intelligence to devices that would be otherwise dumb, enabling them to communicate real-time data without involving a human being. It is making the fabric of the world around us smarter and more responsive, merging the digital and physical universes. It is difficult to comprehend how much data are being generated every day through the use of internet, social media, commercial transactions, digital images, records of health services, government offices, astronomical tracking, emails, security devices, satellite activities, research laboratories, communications, transport, bank cards, weather indices etc. and the list goes on and on. Imagine how much data is produced and processed by Facebook or Instagram or Twitter or any other social media. What about Google, YouTube, LinkedIn, Research Gate, etc.? In a recent study, it was reported that 90% of the entirety of the world's data has been created within the previous two years. In just two years, the world have collected and processed 9x the amount of information than the previous 92,000 years of humankind combined. And it isn't slowing down. Data are now measure in Terabytes, Exabyte, Petabytes, Zettabytes, and Yottabytes etc. Beyond any doubts, the whole world is moving fast into fourth industrial era of big data, artificial intelligence, machine learning technologies and data science to benefit from it. Consequently, the use of unprecedented volume and intensity of big data is becoming more and more an integral part of everyday life of modern science and citizens.

2.1 What are big data?

The rapid development in Information and Communications Technology (ICT) has enabled information to be generated and shared quickly nowadays. Electronic gadgets, such as cellular phones, satellites, Global Positioning Systems (GPS), and scanning devices, and fora like social media and e-commerce create volumes of data on a daily basis, and in some instances by the second. There are over 7 billion mobile phone subscriptions and 3 billion internet users worldwide. Mobile broadband subscriptions increased from 268 million in 2007 to over 2.1 billion in 2013 (ITU, 2013). The information generated by these media constitutes *data exhaust* and is defined as ―the digitally track able or storable actions, choices, and preferences that people generate as they go about their daily lives (Global Pulse, 2012). IBM reported that over 2.5 quintillion bytes of data are generated daily. The stock of digital data rose from 150 Exabyte

in 2005 to 1200 Exabyte in 2010 (Global Pulse, 2012). This kind of fast Moving, high volume data have been dubbed Big Data. The high volume, high velocity and wide variety of these data are commonly referred to as big data (Popoola and Nuhamna, 2019). Big data are data sets that are so large and complex that traditional data-processing applications become insufficient to capture, store, and analyze. Instead, a network of human skills, advanced technologies, and data access infrastructure are essential to handle it. Big data are generated through internet of things. The types, quantity and value of big data are vast: from personal profiles on sites like Facebook or Instagram to demographic data, from bank accounts to medical records to employment profiles. Our web searches and sites visited, including our likes and dislikes and purchase histories; our heart rates, food intake, home temperatures, whether our lights are on or off. The list continues to grow. Big data constitute a source of information that cannot be ignored by official statisticians and as the whole world moves into the fourth industrial era of artificial intelligence and machine learning technologies fully, to benefit from big data official statisticians must brief up. The use of unprecedented volume and intensity of data is becoming more and more an integral part of everyday life of modern science and citizens. The reality is that the fourth industrial era is digital technologies era, it is a data driven, internet dependent and satellite guided era. Starting from driver-less vehicles to store-less shopping platforms and delivery of personalized services will be digitized based on data guided evidence. In fact, big data revolution is already here, and the world is increasingly being exposed to various technologies that are dependent on official data. Big data is continuously produced, generated, and collected electronically with a considerably lower burden on respondents. It can complement traditional statistics with more granularities or, in some cases, even replace traditional data collection methods. With an invaluable continuous flow of digital information about people activities and their impact on society, the economy, and the environment, big data holds tremendous potential for official statistics. Big data permeates many aspects of people's lives, from daily communication and interactions, to shopping and consumption and the medical treatments they receive. Big data could be used to transform the way people and businesses make decisions and measure things. It provides timely, frequent, and granular insights - crucial attributes in critical situations. Big data can also help to monitor the Sustainable Development Goals (SDG) indicators, especially where traditional data are missing. Whereas half of the SDG indicators have no or insufficient data to measure progress in Africa sub- region, big data can address some important gaps. While big data mostly remains at

an experimental phase in Asia and the Pacific, some statistical offices are integrating certain new data sources into the production of official statistics. Fourth industrial era demands a new way of thinking for the production of official data. Indeed, rethinking the central importance of the fundamental of official statistics. Official statisticians may have to accept a trade-off where they sacrifice some aspects of personal privacy within carefully agreed parameters in order to benefit from the collective gains of big data in this digital era.

2.2 Sources of Big Data:

Unlike traditional data sources (sample survey, censuses and administrative) that are compiled for specific purposes, big data is a byproduct "found" in business and administrative systems, social networks, and the internet of things. Social networks are online platforms that help people build social relations with others having similar interests e.g Facebook, Twitter, LinkedIn etc. Users create blogs and profiles, share pictures, and exchange messages and thereby provide human-sourced information that is digitalized and stored. Data in social networks are often ungoverned and unstructured. In its big data classification, the United Nations Economic Commission for Europe (UNECE, 2013) also includes in social networks internet searches and mobile data that can be more widely understood as human-sourced information. Traditional business systems are processes and procedures defined by businesses to provide value to their customers and generate process-mediated data, including administrative records. Business systems record well-governed, structured information on transactions, positions, and metadata related to business events (commercial transactions such as registering a customer or receiving an order) stored in relational database systems. The internet of things is a system of data-producing interrelated computing devices with embedded sensors and internet connectivity that measure and record events and situations in the physical world. Their output is structured machine-generated data (sensor records, computer logs, webcam, and mobile phone location GPS (UNGP. 2012).

2.3 Big Data Era and Official Statistics

Countries in the world have one or more government agencies (e.g national institutes) that feed decision-makers and other users with a steady flow of information. This bulk of data is referred to as official statistics. Official statistics should be objective and easily accessible and produced on a regular basis. Official statistics are generated from the collection and processing of data into statistical information by a government institution or international organization, following and in

line with the principles of the United Nations fundamental principles of official statistics (UNFOPS, 2016). These bulks of data are then disseminated to help users develop their understanding about a particular topic or geographical area, make comparisons between countries or understand changes over time. Official statistics help decision makers develop informed policies that impact millions of people; It also provides a picture of a country or different phenomena through data, and images; It provides basic information for evaluations and assessments at different levels; Official statistics produce relevant, objective and accurate statistics to keep users well informed and assist good policy and decision-making. The advent of Big Data has given rise to data science and this is expected to have a big impact on official statistics and the way in which official data will be procured and analysis to bring about informed decision by the national statistical organizations (NSOs) and national statistical institutes (NSIs). NSOs and NSIs are responsible for official statistics, which are heavily used by policy-makers and other important players in society. Arguably, the way NSIs take up Big Data will eventually have implications for all of society. Official statistics play a key role in modern society. NSIs aim at providing information on all important aspects of society in an impartial way, and according to the highest scientific standards. Information that fulfills these demands is used in public discussion, forms the basis of policy decisions, is required for business use, feeds scientific research, and is used in education and so on. Official statistics can only meet this demand if professional standards play a vital role in securing trust in official statistics. Official statisticians have their own ethics code (United Nations, 2013), which includes an absolute respect for the confidentiality of data provided by respondents. Data collected for statistical purposes may never be disclosed and may never be used for other purposes. At the level of the European Union (EU), quality norms have been codified in the so-called Statistics Code of Practice (Eurostat, 2014). The trust earned by respecting professional standards is also the basis for a privileged position of NSIs in respect of data acquisition. Many NSIs have access by law to government data sources and have the power to collect data from other parties, often without having to pay the provider. Moreover, for statistical purposes, many NSIs are allowed to link data from different sources.

Given the role for NSIs, what does the emergence of Big Data mean for official statistics? This question is addressed in this contribution, but as we will see, there are many reasons why the role

of NSIs in the Big Data era is not 'given'. In order to keep a sound and trusted basis of information for society to rely on, we argue that NSIs may have to adapt to the changing context in which they operate and they can be trusted. In developing countries, official statistics are often taken for granted, but where trust is lacking, society misses an important pillar for informed discussion and evidence-based policy-making.

2.4 Fourth Industrial Era and Official Statistics

"We stand on the brink of a technological revolution that will fundamentally alter the way we live, work, and relate to one another. In its scale, scope, and complexity, the transformation will be unlike anything humankind has experienced before. We do not yet know just how it will unfold, but one thing is clear: the response to it must be integrated and comprehensive, involving all stakeholders of the global polity, from the public and private sectors to academic and civil society" (Schwab, 2015)

In respect of information, society is changing rapidly. For example, there is an enormous growth of data that is gathered and recorded in myriad ways: from satellite and sensory data, to social network and transactional data and so on. The availability of data is also expanding and becoming the foundation of business models. Information is becoming more visual and interactive. Information and communication technology is becoming ever more advanced, processing power and data storage capacity is continuously rising, cloud solutions are emerging and applications are becoming more intelligent. These developments have been described in more depth and detail by many observers, such as Mayer-Schönberger and Cukier (2013).

These changes have many impacts on societies. For one, the increased gathering of data and the commercial and social possibilities of data usage influence public opinion on privacy. Some are concerned if their data are re-used without their consent, for commercial reasons or otherwise. Others do not mind so much, if this means that services are provided for free. Many people voluntarily share information on social networks without caring for privacy. People have less patience to fill in questionnaires, especially if the data requested have been registered somewhere else already. Government agencies are expected to be more forthcoming in providing data. Governments have reacted to the changes by formulating policies on, for instance, open data and availability of public sector information, also at the EU level (European Union, 2013).

How have NSIs responded? Until around the 1980s, data were essentially a scarce commodity with a high price. Before the era of Big Data, information was not readily available but had to be collected for a particular purpose. Official statistical information based on survey data had a unique value: there simply was no alternative. For example, population census data, collected door to door, was immensely valuable to policy-makers, researchers and other users. In the last few decades, data collected by public administrations have become increasingly accessible for statistical purposes, stimulated in part by information technology developments. Statistical data collection by means of questionnaires was supplemented and increasingly replaced by administrative data sources. Nowadays, some countries do not conduct extensive population surveys anymore but compile census statistics by combining and analyzing data from several administrative sources. NSIs became more integrated in the information architecture of the government. In this way, the burden on persons and businesses to respond to questionnaires was considerably reduced.

In the context of all of these developments, the information provided by NSIs still remained unique. In particular, the possibility of combining data from different sources made official statistics even more valuable, since in many countries no other organization was positioned to do so. In parallel, efforts also increased to standardize and harmonize these various sources of official statistics, especially in the European Union. Supported by legislation, official statistics in the European Union are now considered a system, the so-called European Statistical System, or ESS (Laureti et.al., 2021). However, Big Data is changing the environment of the NSIs once more as data scarcity is becoming less of an issue. For NSIs, there are potential benefits as new data sources and opportunities emerge. But it also makes the products of NSIs potentially less unique, since other players in the information market may start – and have actually started – producing statistics, for instance, on inflation, such as the Billion Prices Project of MIT (Ricciato et.al.,2019).

2.5 Big Data: Opportunity for Official Statistics

Let us first look at the opportunities for NSIs offered by Big Data. There is a huge potential for new statistics (Daas et al., 2013). Location data for mobile phones could be used for almost instantaneous daytime population and tourism statistics (De Jonge et al., 2012). Social media

messages could be used for several types of indicators, such as an early indicator of consumer confidence. Inflation figures could be derived from price information on the web, and so on. In addition, Big Data sources may be used to substitute or supplement more traditional data sources, such as questionnaire and administrative data. For instance, data collection by questionnaire on road use may not be necessary anymore if detailed traffic loop data, i.e. data from sensors in roads, become available (Struijs and Daas, 2013).

3.0 Emergence of Data science:

The limitation of conventional statistics to manage and analyze big data has inspired data analysts to venture into data science. Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data, (*Dhar, 2013*), (Jeff, 2013) and apply knowledge and actionable insights from data across a broad range of application domains. Data science is related to data mining, machine learning and big data. Data Science is a combination of multiple disciplines that use statistics, data analysis, and machine learning to analyze and extracting meaningful insights from the complex and large sets of data (big data). It is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the complex and large sets of data. Data science is related to data mining, machine learning and big data. Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" in order to "understand and analyze actual phenomena" with data (*Hayashi,1998)* It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge (*Cao, 2017)* However, data science is different from computer science and information science. According to Tony and Bell (2009), defined data science as a "fourth paradigm" of science (empirical, theoretical, computational, and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge. Data science not only requires conventional statistical methods, it also needs skills such as statistical signal processing, pattern recognition, data mining, machine learning, bioinformatics, meta-analysis etc. (Khan, 2020). Khan (2020) discussed various statistical models to Meta-analyse data from heterogeneous primary studies including the inverse variance heterogeneity (IVhet) model. Unlike the conventional data analyses which is easily managed by personal computers or laptops, the storage and analysis of

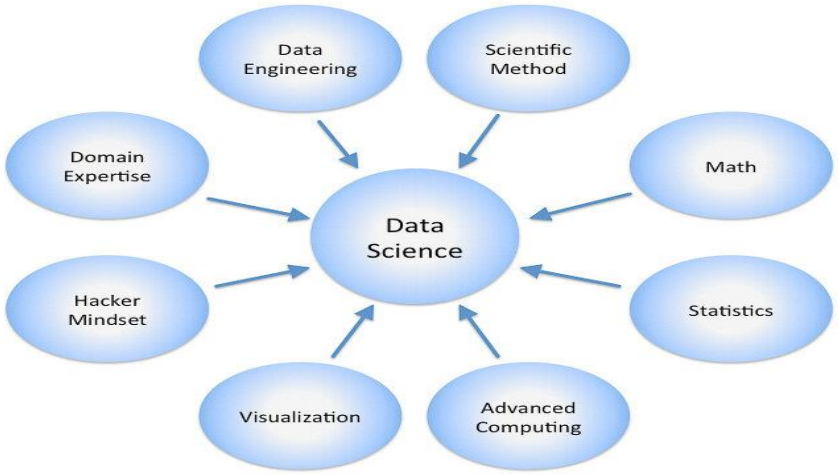
big data requires cloud technology. Sahinoglu and Cueva-Parra (2010) Cloud computing is the use of computing resources (hardware and software) that are delivered as a service typically over the Internet. The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software & computation.

3.1 How Does Data Science Work?

Data science involves a plethora of disciplines and expertise areas to produce a holistic, thorough and refined look into raw data. Data scientists must be skilled in everything from data engineering, math, statistics, advanced computing and visualizations to be able to effectively sift through muddled masses of information and communicate only the most vital bits that will help drive innovation and efficiency. Data scientists also rely heavily on artificial intelligence, especially its subfields of machine learning and deep learning, to create models and make predictions using algorithms and other techniques.

Here is a diagram showing some of the common disciplines that a data scientist may draw upon. A data scientist’s level of experience and knowledge in each often varies along a scale ranging from beginner, to proficient, and to expert, in the ideal case.

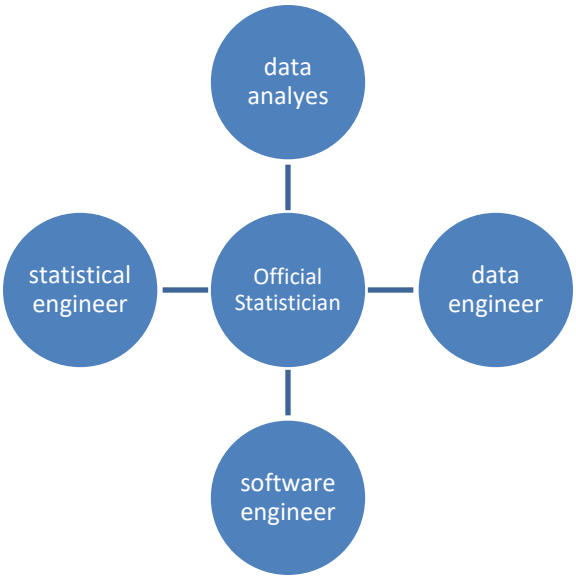
Figure 1.1: The General Data Science Cycle



3.3 Data Science and Official Statistics

A profound disruptive digital transformation in society and economy is taking place worldwide, with governments, private enterprises, and the public demanding quicker access to real-time quality data for better decision-making and social good, an output that often goes well beyond traditional statistical production capabilities. The life cycle of traditional statistical information is becoming increasingly shorter. The longer an end-user must wait for statistical information, the less valuable it becomes for decision-making. Hence, the ability to incorporate new and Big Data sources and to benefit from emerging technologies such as Web Technologies, Remote Data Collection methods, User Experience Platforms, and Trusted Smart Statistics will become increasingly important in producing and disseminating official statistics in this era. The skills and competencies required to automate, analyze, and optimize such complex systems are often not part of the traditional skill of most National Statistical Offices. The adoption of these technologies requires new knowledge, methodologies and the upgrading of the quality assurance framework, technology, security, privacy, and legal matters. However, there are methodological challenges and discussions among scholars about the diverse methodical confinement and the wide array of skills and competencies considered relevant for those working with big data at NSOs. For official statisticians to benefit from data science era, a combination of statistical engineers, data engineers, data analysis, software engineers and soft skills such as statistical thinking, statistical literacy and specific knowledge of official statistics and dissemination of official statistics products as key requirements. The Data science diagram in figure 1.1 will change to the below diagram

Figure 1.2: Data Science in Official Statistics Cycle



4.0 Challenges of Data science for official statistics

No doubt, the world is moving fully into the fourth industrial era of data science in which swathes of new digital data are valuable for official statistics. The questions to ask are these: how would official statistics benefit from it and what are the possible challenges for official statistics?

The use of novel statistical techniques (e.g., machine learning, artificial intelligence, natural language processing) approaches to get additional insight and understanding in every field of human activity from analyzing traditional data (e.g., administrative databases) and new sources of data (big data) collected from digital activity e.g web activity, social networks, online payments, transportation systems – mobile phone usage, remote sensing networks, satellite imagery, Internet of Things (machine-generated data) and to build tools to inform decision-making has become a crucial challenge (and opportunity) for national statistical offices (NSOs). There is growing demands for trusted information, fast-development (achieving development goals agenda and Africa Agenda 2062) in all the regions of the world. NSOs and NSIs need to take the advantage of data science technologies and accessible technologies in collecting, analyzing, and disseminating more frequent and timely information for the decision makers with a reduced burden on respondents. NSOs NSIs need to take advantage of the potential of mobile technologies, machine learning, augmented intelligence and fast-growing (including cloud)

computing capabilities to provide timely data for decision makers. These new approaches are not expected to replace traditional methodologies at NSOs. Contrariwise, they have the potential to become an important complement to official statistics in meeting their objectives of providing timely and accurate evidence for public and private decision-making. New data sources and data science techniques have the potential to empower national statistical systems (NSSs). For NSOs and NSIs to embrace data science technologies and methodologies for official statistics, the following challenges must be overcome.

4.1 **Capacity Building for Official Statistician:** Official statistics have sets of rule for their various operations: from data collection, data codding to data analysis; data analysis to information dissemination and to data communications. Creating, up skilling, developing and retaining the set of skills knowledge and talent required to extract all the potential from using Big Data acquisition, processing, analysis and visualization in the statistical production process are critical challenges NSOs will have to quickly address. This will have to be done both at the organizational level, identifying strengths and weaknesses, setting goals, establishing a roadmap and a strategic plan for training programs, and at the individual level, identifying gaps in core and soft skills and competencies, updating personal development plans. There is need to build the capacity of staff on the collection of big data (Statistical Engineer), big data storage (data mining), big data analysis (Statistical Analyst), data Visualization and competencies at NSOs. Appropriate knowledge of high-frequency data, spatial data, big data, micro-data are considered important for the future of data stewardship in official Statistics, which would support the value-added of new data infrastructures. Developing core competencies in the new data methodologies and in the new data quality tools and frameworks are critical for the success of transformation of official statistics. Regarding general (soft) skills, empowering problem-solving skills, data and statistical literacy, and all indicators of ethics are considered important for the adoption of data science competencies for official statistics production process.

4.2 **Flexibility in NSOs Operations**: Major technological achievements may imply significant public policy issues. McKinsey (2016) in its report underlines that the key for the successful adaption to the new technological conditions is the ability of NSOs to adopt the right policies. National Statistical Organizations that will not be able to follow the appropriate long-term policies will set their nations at risk, that is, when all the other NSOs will run with great speed,

their inability to be adapted to the new reality will drive to the deterioration of their relevancy, the reduction of their values, and the increase in their spending with the possibility of a bankruptcy to be increased. Given that the fourth industrial era is directly related to socioeconomic growth, these policies must be in complete accordance to the Sustainable Development Goals (SGs) adopted by United Nations Member States in 2015 (Smit et. al, 2016), (Arntz et.al., 2016) But it is not only the ability of NSOs to be adapted to the new conditions. There are also severe social problems that may get bigger due to the fourth industrial era making policy intervention crucial. Political leaders must ensure that the technological progress will work for the benefit of NSOs and not against it.

4.3 **Provision of data processing and smart technologies:** Machine learning and data visualization novel tools have already started to serve official statistics products (Zacardi and Infante, 2021). However, the future usage of these two concepts is expected to experience a rapid increase. The use of machine learning and data visualization tools is projected to growth exponentially in preparing official statistics. The empirical results also show that the massive use of traditional software such as SAS might decrease and the popularity of open-source programming languages such as R software are expected to increase at NSOs. The results also suggest that regardless of the positive benefits of data science for official statistics, there is a risk of less transparency in official statistics in future.

4.4 **Building appropriate information and communications technology infrastructure for data storage**: NSOs will be called to identify and assess new data sources, to set up new partnerships and collaboration agreements with multiple stakeholders. They will need to build appropriate information and communications technology infrastructure for data storage (offsite and onsite) and analytics, including processing power, integrating files from numerous sources different formats, arriving potentially at different times with a different degree of reliability.

4.5 **Review of legislative, ethical and data security issues:** The new data sources and data science techniques have the potential to empower national statistical systems. NSOs need to revisit old and new legal and ethical dilemmas on data access and sharing (privacy and data protection), particularly outside government data. The right to access administrative data, established in principle by the law, is not adequately supported by specific obligations for big data. Many potential Big Data sources are collected by non-governmental organizations or are

freely available on the web; situations that may not be covered by existing legislation, hence, for NSOs to benefit greatly in this era, the whole ethnic and legal issues need to be revisited.

5.0 Conclusion:

Major waves of technological progress such that of Fourth Industrial Revolution always create concerns about the future of human labor and the possibility of substitution of the human factor by machines and robots. Also, the rapid rate of technological change and commercialization in using digital data is undermining confidence and trust. Tensions are rising. Concerns about the misuse of digital data continue to grow. Also, mounting is a general public unease about what "they" know about us, as confirmed by the Snowden revelations. Fundamental questions about privacy, property, global governance, human rights – essentially around who should benefit from the products and services built upon digital data — are major uncertainties shaping the opportunities. National Statistical Organizations can benefit greatly if all these concerns are looked into and find ways to answer to some of the fundamental questions raised.

References:

1. Arntz M, Gregory T, Zierahn U.(2016). The risk of automation for jobs in OECD countries: A comparative analysis. In: OECD Social, Employment and Migration Working Papers, No. 189
2. Bell, G.; Hey, T.; Szalay, A. (2009). "Computer Science: Beyond the Data Deluge". Science. 1297–1298.
3. Cao, Longbing (2017). "Data Science: A Comprehensive Overview". ACM Computing Surveys. **50** (3): 43:1–43:42.
4. Daas PJH, Puts MJ, Buelens B, et al. (2013) Big Data and official statistics. Paper for the 2013 NTTS conference, Brussels, Belgium, 5–7.
5. Dan Miller. (2015). Natural Language: The User Interface for the Fourth Industrial Revolution
6. De Jonge E, Van Pelt M and Roos M (2012) Time patterns, geospatial clustering and mobility statistics based on mobile phone network data. Discussion paper, Statistics Netherlands
7. Dhar, V. (2013). "Data science and prediction". Communications of the ACM. **56** (12): 64–73.
8. Eurostat (2014). European statistics code of practice
9. European Union (2013).

10. Global Pulse (2012) Big Data for Development: Challenges & Opportunities
11. Hayashi, Chikio (1998). "What is Data Science? Fundamental Concepts and a Heuristic Example". Data Analysis, and Knowledge Organization. Springer Japan. pp. 40–51.
12. ITU (2013). Big Data: Big Today, Normal Tomorrow
13. Jeff, Leek (2013). 'The key word in "Data Science" is not Data, it is Science". Simply Statistics.
14. Khan, S. (2020). Meta-Analysis Methods for Health and Experimental Studies. Singapore: Springer Nature.
15. Laureti T, Benedetti I, Palumbo L, Rose B.(2021). Computation of consumer spatial price indexes over time using Natural Language Processing and web scraping techniques.
16. Mayer-Schönberger, V, Cukier, K (2013) Big Data: A Revolution that Will Transform How We Live, Work, and Think, London: John Murray Publishers.
17. Osuolale P. Popoola and Nicholar N. Nuamah (2018)."New Trend in Modelling Climate Change in the Era of Big Data" Anale. Seria Informatică. Vol. XVI fasc. 2 – 2018.
18. Ricciato F, Wirthmann A, Giannakouris K, Reis And F, Skaliotis M.(2019). Trusted smart statistics: Motivations and principles. In: Statistical Journal of the IAOS. IOS Press; pp. 589–603.
19. Sahinoglu, M. and Cueva-Parra, L. (2011). CLOUD computing. Wiley Interdisciplinary Reviews: Computational Statistics, 3(1), 47-68.
20. Schwab, K. (2015). The Fourth Industrial Revolution: What It Means and How to Respond
21. Smit J, Kreutzer S, Moeller C, Carlberg M.(2016). Directorate General for Internal Policies Policy Department A: Economic and Scientific Policy Industry 4.0. European Union
22. Struijs P and Daas PJH (2013). Big Data, big impact? Paper presented at the seminar on statistical data collection, Geneva, Switzerland, 25–27.
23. UNECE (2013a). What does 'Big Data' mean for official statistics? Paper prepared on behalf of the high-level group for the modernisation of statistical production and services,
24. UNECE (2013b). The role of Big Data in the modernisation of statistical production. Project plan.
25. United Nations (2013). Fundamental principles of official statistics
26. United Nations Economic Commission for Europe (2013). Classification of Types of Big Data.

27. United Nations Global Pulse (2012). “Big Data for Development: Challenges and Opportunities
28. Zaccardi J, Infante E.(2021). A systematic approach for data validation using data driven visualisations and interactive reporting. In: Conference on New Techniques and Technologies for Official Statistics.