The Use of Big Data in the Field of Official Statistics
(Descriptive Exploratory Study)

Researcher / Munthir Mohammad Al-Ansari
Senior Consultant
Tourism Business Intelligence Assistance Agency
Ministry of Tourism, Kingdom of Saudi Arabia
September 01, 2022

Abstract:

This study aims to reveal the mechanisms and role of big data in the development of official statistics. data “, while traditional methods of producing official statistics in terms of data sources depend mainly on surveys and administrative records data and their delays in meeting the needs of beneficiaries of official statistics, access to data produced by our electronic devices as well as by the Internet and through social networks has emerged. In order to achieve the study's objective, previous studies and the experiences of some States and organizations in the use of big data for the development of official statistics have been surveyed using the descriptive research curriculum through previous studies and literature that enables the researcher to access them.

The study concluded the importance of building the capacity of specialized personnel in various areas of information technology; to assist in analyzing big data and producing official statistics. and the need to raise awareness of the importance of big data among those working in the official statistical sector from leaders and employees in public and private institutions. Also, work to build a national strategy for the development of official statistics using non-traditional data sources, including big data. The data available to various government and private entities and social media are invested through pilot projects to produce official statistics. The study also emphasized the need for a national governance framework to ensure the security and protection of statistical items' data privacy.

Keywords: big data, official statistics, IT.

First Topic: Methodological Framework of the Study

Over the past few years, the world has witnessed a great development in information and communication technology, and a huge increase in the amount of data that is produced, stored and made available through networks, which has resulted in an inflation of digital data available through various communication channels. Data that was previously seen as unimportant can now add up to critical information when combined with data from other sources. It has become possible to process large amounts of diverse data at high speed in real time. The volume of data available around the world has increased due to the activity of the digital revolution through smart devices, which is what specialists refer to as “Big Data Deluge” (Maqnani and Moqaddam, 2019).

The traditional methods of producing official statistics in terms of data sources depend mainly on survey data and administrative records data. However, in our modern world, data generated from the Internet and through social networks and others is increasing rapidly and steadily, as well as the huge amount of data produced by the electronic devices surrounding us. Big data tools became a widespread solution. Not only are they used in specialized IT industries, but in all cases where there are data handling requirements that cannot be handled by traditional IT solutions.

At present, the United Nations (UN) and the United Nations Economic Commission for Europe (UNECE) as well as several National Statistics Offices (NSO) are working on pilot projects aimed at exploring possible uses of big data in the production of official statistics. This global movement aims to reach unified and agreed methodologies, define technical requirements and establish a legislative framework for accessing big data sources and determining the required knowledge and skills.

In light of the above, the current research seeks to study the role of big data in developing official statistics by exploring previous studies and the experiences of some countries and organizations in using big data for the purpose of official statistics.

The Study Problem:

Big data differs in nature from traditional sources of data in terms of:

1. The huge and massive volume of this data (Volume).
2. The rapid, continuous and real-time production and dissemination of this data (Velocity).
3. The diversity of this data between structured, unstructured and semi-structured (Variety).

The problem of the study is to identify the possibility and areas of using big data in official statistics, as it represents an updated source of data that can be used in official statistics.

Objectives of the Study:

1. Exploring the role of big data in the production of official statistical data.
2. Learning about the techniques used in big data management.
3. Learning about ways to invest in big data and statistical documents.
4. Determining the applications of using big data in the production of official statistics.
5. Determining the challenges of using big data in national centers and authorities for official statistics.

Significance of the Study:

The significance of the study is to explore the potential applications of big data sources, which are among the non-traditional data sources by taking advantage of the huge volume of data resulting from the digital revolution and its competition with traditional sources of official statistics. The traditional methods of producing official statistics depend mainly on survey data and administrative records data, as the only sources of data in official statistics at the global level and in the Kingdom of Saudi Arabia. However, in our modern world, data generated from the Internet and through social networks and others is increasing rapidly and steadily, as well as the huge amount of data produced by the electronic devices surrounding us.

Methodology of the Study:

Based on the problem of the study, its question, and its objectives, the approach used in this study is a descriptive research approach through previous studies and literature that the researcher was able to view.

The Second Topic: Theoretical Framework of the Study

First: Introduction to Big Data?

Watson (2014) defined big data as "a variety of data in large quantities that are difficult to handle well", and Bieraugel (2016) defined it as "data that cannot be stored or analyzed by traditional hardware and software". As for the procedural definition of big data, it is that "large-sized data that is stored in an unorganized manner that does not make it easy to benefit from it" (Al-Aklabi, 2017).

1. Definition of Big Data

Canopy (2015) defines big data as "a term used to describe sets of data that are extremely large or of a high degree of complexity, or that require a great deal of rapid processing - sometimes called volume/diversity/speed problems, that become difficult or impossible to address using traditional databases and traditional analytical tools, knowing that processing data of this size requires software that runs in parallel on dozens, hundreds, or even thousands of servers" (Canopy, 2015).

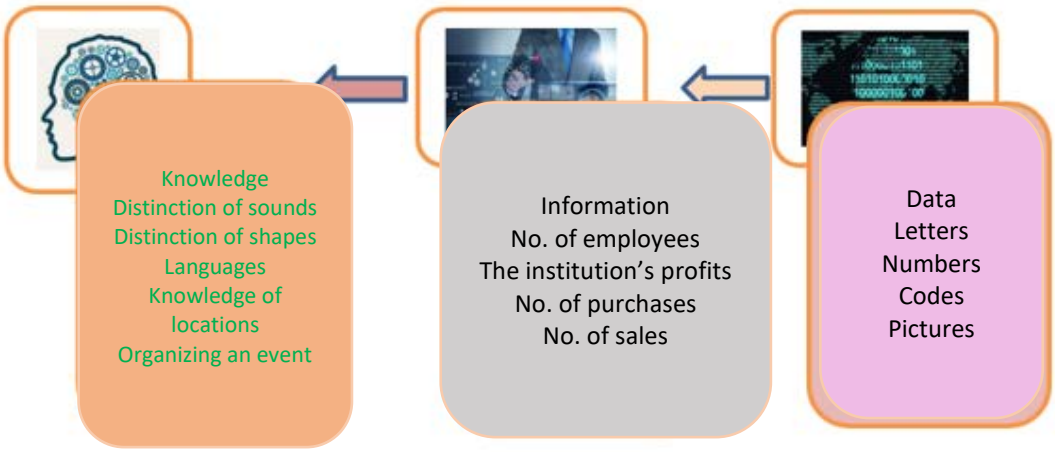
Big data is also defined as:
"A wide and very complex set of data with its own characteristics (such as size, speed, diversity, variance, validity of the data), it is difficult to process it using current technology to make use of it. The challenges that accompany this type of data lie in processing, providing, analyzing, storing, searching, sharing, transferring, photographing and updating it, as well as preserving the privacy that accompanies it" (Al-Bar, 20147).

Based on the aforesaid definitions, big data can be defined as: a huge variety of multiform data (read, audio, visual, etc.) produced by various sources. In order to benefit from it, it should be addressed through non-traditional technological means, that is, this term refers to the inflated volume of data in terms of its number, speed, and diversity in its production. This data comes from digital photos and videos, social networks, smart devices, and smart cell phones.

2. The Difference between Data, Information and Knowledge

Data can be defined as a set of letters, words, numbers, symbols, or pictures related to a topic. Data by itself has no meaning or value, and it is the raw image of information, such as employee data and their photos (Sanders, 2016, pp. 1-3). As for information, it is data that has been analyzed and processed, so that it carries meaning and value, and can be used in decision-making, such as obtaining the total number of employees, their average age, average years of experience, and others, while knowledge is the process of analyzing different information and linking it to each other and having a clear understanding of it and mixing it with experience.

Figure (1) The difference between data, information and knowledge:



Source: Abstract adapted from several sources and published studies

3. Types of Big Data

Big data can be divided into three types (Meguneni & Mokedem, 2018):

A. **Structured Data**: It is the data organized in the form of tables or databases in preparation for processing it.
B. **Unstructured Data**: It constitutes the largest percentage of data, and it is the data that people generate daily from text writings, images, videos, messages, clicks on websites, and others.
C. **Semi-structured Data**: It is a type of structured data, but it is not designed into tables or databases.

4. Big Data Sources

The sources of big data can be identified according to the following sources (Hamlawi, 2018):

A. **First Source: Data Produced by Institutions**: Data that is produced by companies and institutions during their work on a daily basis, and is stored in databases or files that are specific to the company or institution.
B. **The Second Source: Human-generated Data**: human-generated data, whether documents on computers or those stored in various mobile devices, such as text messages and smart device applications, in addition to that data that people include in the global network of information, like email and social media, from sharing photos, conversations, and comments, much of it is unstructured data.
C. **The Third Source: Data Produced by Machines**: It is all data that is produced automatically without human intervention, such as data from surveillance cameras, sensors, and satellites.
D. **The Fourth Source: Data Tracking Devices**: such as tracking data derived from mobile phones and GPS.

5. Other Sources of Big Data

A. Social Media Sources

Statistics published by the Ministry of Communications and Information Technology in March 2016 show that there are 18.3 million Saudi users subscribed to social media platforms, with a usage rate of 58% of the total population. The average time an individual spends on social media is 260 minutes for those who access social networks through smartphones (i.e. more than 4 hours a day). Social data can be thought of as a large discussion group that generates rich and detailed data, which can be used to measure many economic and social statistics. Examples of new statistical products that can be produced based on social networks are:

- Consumer Confidence Index.
- Subjective Well-being Index.
- National Cohesion Index.
- happiness index.

The above products fall under the category of new indicators that can be calculated based on big data. The subjective well-being index is known as an important indicator that outperforms monetary measures (such as GDP) in measuring people's well-being. For its part, the Organization for Economic Co-operation and Development developed guidelines on how to objectively measure well-being using survey studies (OECD).

B. Mobile Data

The statistics of the Ministry of Communications and Information Technology revealed that the total number of mobile phone subscribers was estimated at 44.5 million by the end of the first quarter of 2017. The level of coverage will rise to 140% of the total population. This provides a great opportunity for the General Authority for Statistics to use such a source in the production of official statistics. For example, tourism statistics, transportation statistics, and population statistics are of interest

C. Point of Sale Data:

There are a good number of supermarket chains with good coverage of food products. The data of the scanners in these stores can be used to calculate the consumer price index. This results in immediate cost reduction by reducing the scanning range for items not in these stores, and this process leads to a more accurate consumer price index. The experience of the Australian Bureau of Statistics can be cited as the largest example of the use of big data. Since March 2014, the Australian Bureau of Statistics has used this type of data to price over 25% of all items used in the consumer price calculation.

D. **Sensors' Data**:

With the spread of sensors on the main roads, it became possible to search for the possibility of using this data to calculate traffic density statistics. The best experience to guide in this field is the experience of the Dutch Statistics Office, which relies entirely on sensor data in calculating road statistics.

Second: Roles and Skills of Using and Managing Big Data

The following table summarizes the different roles required to use and manage a big data architecture. Each role requires a different set of skills, which are listed in the following table:

Table (1) The roles and Skills of those Responsible for Big Data Management:

Role	Function	Skills
Directors	Creation and management of structure in terms of hardware and software	Advanced Linux skills required. Networking skills are required only if there are clusters of physical nodes, you need to develop specific skills for Hadoop configuration and management. This role is not required in the case of cloud deployment
Programmers	Developing Spark programs	Python programming and developing specific Spark programming skills
Statistical Analysts	Analyzing data sets and giving specifications to programmers about statistical methods	General statistics, knowledge of database language for self-service analysis is recommended.
Data Scientists	Data analysis using advanced analytics techniques	Python programming, machine learning algorithms.

1. **Big Data Formats**

The big data is as follows (Al-Tayyib and Al-Rubai, 2018):

A. **Web Data**: It can capture individual-level web business data such as page views, searches, and reading comments. It can enhance performance in areas such as best offers, building forms, division of individuals and targeted ads.

B. **Text data**: email, news, Facebook posts, documents, etc. It is one of the largest and most widely applied types of data.

C. **Time and location data**: GPS and mobile phone technologies, along with time and location information, make it a big data provider, and should be treated with extreme privacy caution.

D. **Social network data**: through web applications 2 through social networks.

Correlation analysis can be done to detect a network from a specific user, where social network analysis can give insights about which ads might attract specific users.

2. Measuring the Size of the Data

The size of the data can be measured using a set of different measurement units that start with the lowest thing, which are bits, bytes, and kilobytes, reaching the highest known unit, which is Domegemegrottebyte (Weiss, A., 2018, p. 19)

The Bit: It is the smallest possible storage unit, representing the passage or non-passage of an electric current, and the value of the bit is 1 or 0, where 1 represents the passage of an electric current, and 0 does not pass (Ward, 2004). The unit of measurement for storage capacity is the byte and its multiples are shown, as shown in the following table.

Table (2) Criteria for Measuring Data Size:

Measurement Unit	Abbreviation	Size
Bit	**B**	0 or 1
Byte	**B**	8 Bit
Kilobyte	**kB**	1024 Byte
Megabyte	**MB**	1024 Kilobyte
Gigabyte	**GB**	1024 Megabyte
Terabyte	**TB**	1024 Gigabyte
Petabyte	**PB**	1024 Terabyte
Exabyte	**EB**	1024 Petabyte
Zettbyte	**ZB**	1024 Exabyte
Yottabyte	**YB**	1024 Zettbyte
Xenottabyte	**XB**	1024 Yottabyte
Shilentnobyte	**SB**	1024 Xenottabyte
Domegemegrottebyte	**DB**	1024 Shilentnobyte

3. Data Storage Methods:

Data storage methods vary between:

A. **Structured data**: It is stored in an organized manner, such as data in databases, and is characterized by ease of processing and analysis.

B. **Unstructured data**: It represents the largest percentage of data, and it is present and stored in an unorganized manner, such as data on the global network of information (the Internet), especially social networks and smart phone applications, and it cannot be easily processed.

C. **Semi-structured data**: stored in a specific format and arrangement different from databases (Ward, 2004).

Figure 2: Data Storage Methods

Structured Data	Unstructured Data	Semi-structured Data
Data organized in a table or database, such as data on employees, bank clients or transactions of various institutions.	It represents the largest percentage of data, such as the data that websites produce daily from texts, images and videos, especially social networks, as well as smart device apps.	It is a type of structured data that is not shown in tables or databases, but is orderly stored in files: JSON, HTML and XML
		

Third: The Importance of Using Big Data

Big data analysis helps to reach a more comprehensive and detailed understanding of the characteristics and needs of individuals and different groups by including multiple and disparate pieces of data in the analysis process. Thus, the services and offers provided to these groups or individuals can be more specialized and appropriate, and communication and interaction with them will be more efficient. As a result, satisfaction among individuals can rise, and big data can enhance the efficiency and accuracy of predictions. Big data can also be used to discover cost-cutting opportunities and efficiency gains through comprehensive analysis of various organizational and operational processes, in addition to productivity gains, and increased access to various forms of innovation (Ministry of Information and Communication Technology, 2014).

Fourth: The Development of Data Processing Systems

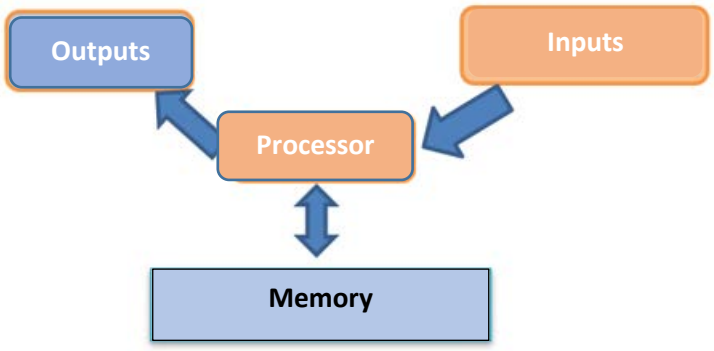
Data processing methods and mechanisms have evolved, as they have gone through several stages in connection with the development witnessed by the structural structure, whether in terms of the development of devices, their components, their technical specifications, and the mechanism of linking them to each other, or according to the development of the operating systems themselves, and this development can be summarized through the most important three stages as follows:

A. The First Stage: (Serial Systems)

With the invention of the first computer in 1945, the devices were mainly based on the implementation of tasks on the use of the Central Processing Unit (CPU), where program instructions and data processing are executed in electronic memory in a sequential manner, hence the name comes from it. Note that this required a very large time to implement the instructions sequentially one after the other (Mercier, 2019).

The following diagram shows how the serial systems work:

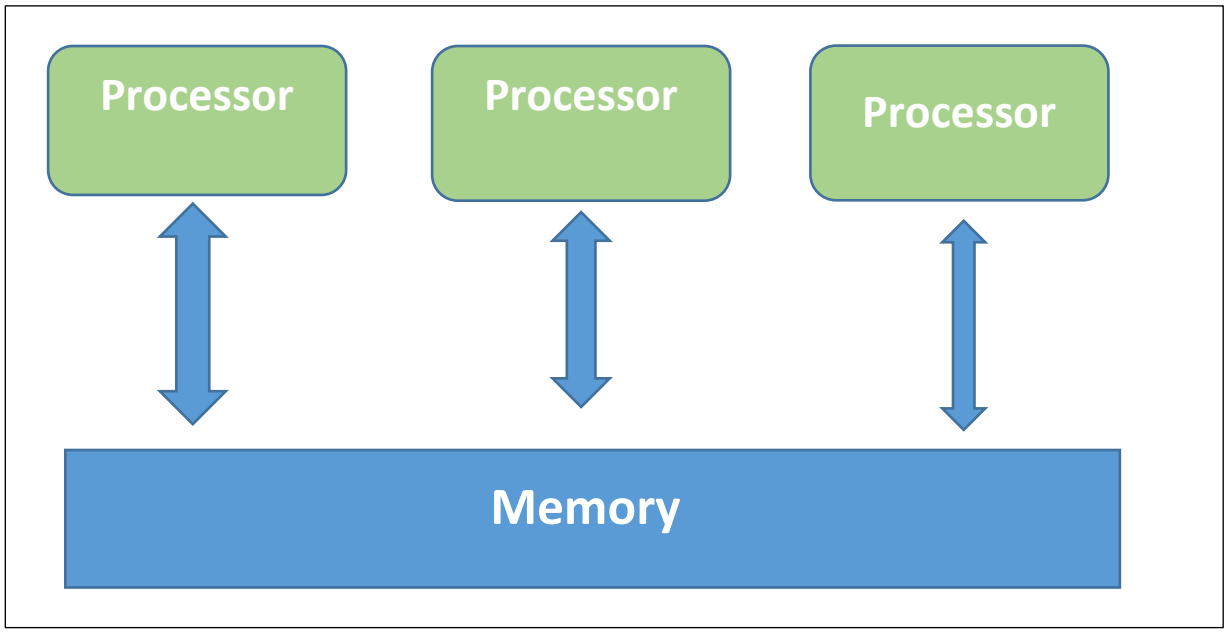
Figure 3: How serial systems work



B. The Second Stage: (Parallel Systems)

Because of the large time required to implement programs and process data, as well as the inability to perform more than one task at the same time, the idea of increasing the efficiency of computers and enhancing their specifications to give them higher performance and improve processing and response time has been a major concern for specialists. In fact, they succeeded in that, by increasing the ability to perform simultaneous performance, that is, the ability of the device to perform more than one task at the same time by increasing the number of processors and the size of memory, which is known as parallel systems, which caused a boom in many fields, especially space and aviation, which can be embodied in the following drawing:

Figure 4: How Parallel Systems Work

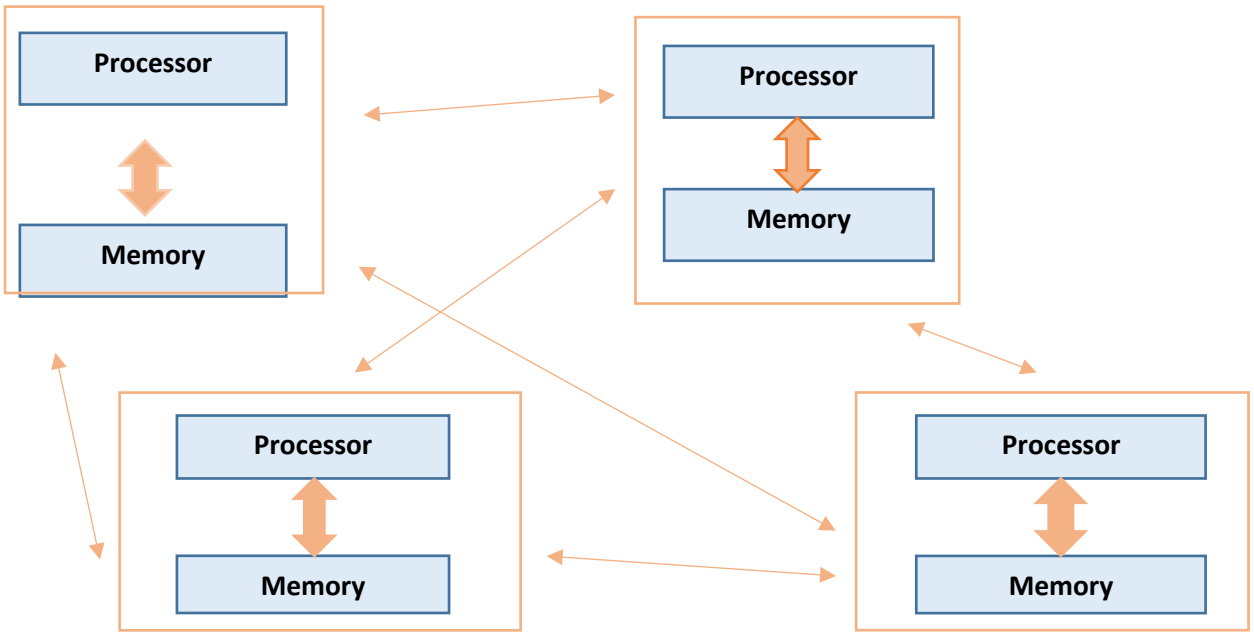


Source:

C. The Third Stage: Distributed System:

With the beginning of the eighties, which witnessed the development of networks and the ability to connect more than one device in one network, a new dimension has appeared in thinking in another direction, which is the implementation of those big tasks which still need a lot of processing time and requires more storage space and memory jointly, synchronously and distributed between more than one device known as distributed systems technology. As the figure below shows, distributed systems are systems that operate on many devices linked together by a network, that is, they are operating systems that appear to the user as if they are running on a central system, but in fact they are running on several central processing units operating at the distribution level.

Figure (5): How Distributed Systems Work



The Third Topic: The Applied Framework of the Study

First: National Statistics

1. Background on National Statistics

The business model of national statistical offices currently relies on two data sources: administrative surveys and records. The development of data collection and storage technology and the proliferation of social media platforms have revolutionized data production, leading to the emergence of big data. This data is characterized by its huge size and real-time in (Velocity) and (Variety).

At present, big data tools are becoming a widespread solution. Not only are they used in specialized IT industries, but in all cases, there are data handling requirements that cannot be handled by traditional IT solutions. The United Nations (UN) and the United Nations Economic Commission for Europe (UNECE) as well as several National Statistics Offices (NSO) are working on empirical projects aimed at exploring the possible uses of big data in the production of official statistics. This global movement aims to reach unified and agreed upon methodologies, determining technical requirements, establishing a legislative framework to access big data sources and setting the required knowledge and skills.

The traditional methods of producing official statistics in terms of data sources depend mainly on survey data and administrative records data. However, in our modern world, the data generated by the Internet and through social networks and others is increasing rapidly and steadily, as well as the huge amount of data produced by the electronic devices surrounding us.

This data differs in nature from traditional sources:

- ✓ The huge and massive size of this data (Volume).
- ✓ The speed of production and dissemination of this data in a rapid, continuous and instantaneous manner (Velocity).
- ✓ The diversity of this data between structured, unstructured and semi-structured (Variety).

From this standpoint, the concept and term “Big Data” arose, and this led to the generation of new business opportunities in the private sector. Big data also represents a very important field as an input in the production of official statistics. Big data provides new and rich sources of data that can be important in the production of statistics, for example:

- ✓ Online data on prices, products, jobs, etc.
- ✓ Data on social networks and the opinions and interests of citizens represented by it.
- ✓ Data on the behavior of citizens on Internet pages by searching for a specific product or treatment for a specific disease.

- ✓ Data generated by electronic devices such as data generated by the use of mobile devices and GPS.
- ✓ Data issued by sensors such as satellite images, traffic lights and climate sensors.

The scope of use of big data can be categorized into three different types:

A. **Generating new statistical products**: In such a case, new products are developed using the big data source, which means that the product cannot be calculated from surveys and administrative records. An example of this is traffic statistics at the road level using camera data.
B. **Complementing the existing statistical products**: In such a case, large data sources are combined with traditional sources, in order to reduce cost or increase accuracy. An example of this is the use of point-of-sale data in shops to calculate part of the prices of items used in calculating the consumer price index (for example, food products), provided that items not available in hypermarkets will be supplemented by surveying.
C. **Replacement of the existing statistical products**: As an extension of the previous use case, data sources can replace traditional sources with big data. For example, statistical offices can produce an index of consumer confidence entirely through the use of social media data.

2. Using Big Data Techniques for Statistics

In this section we will discuss the specialized use of Big Data Technology within official statistical agencies. Beginning with an overview of potential uses, we will review international experiences and provide lessons learned from them.

We can perceive four possible uses:

A. Big data collection: Data is collected as it is directly in the big database, where it is pre-processed for refinement and/or sampling. The revised data is then stored and extracted in an RDBMS, where it can be distributed to users and/or processed more broadly. The advantage of using big data tools at an early stage of processing is that they can handle very well the large volume of data from big data sources. This allows both the original data and the differently processed versions to be preserved without having to delete the data in order to save space. Analysts can return to the original data source at any time to test different solutions or use the same set of data for different purposes.
B. **Database dump**: Relational DBMS stores and processes data as needed. When the volume becomes critically redundant, big data tools make the loading of data into databases, for example storing all historical data. This solution is suitable for cases where the big data sources are for a statistical process that iterates over time, with ever-growing data. It allows keeping the space used for data storage and database performance under control, after

storing data in databases it is always available for analysis at any time without the need to connect to the network.

C. **High-density Processing**: Big data tools can host multiple copies of datasets that are not “big data” in nature, but are large enough to cause problems in traditional tools when subjected to complex processing. As an example, we can look at the problem of linking records, where we must link a data set representing a population (of individuals or institutions) to other data sets and a large number of comparisons. In such cases, big data tools can take advantage of symmetry to achieve fast performance and to be able to also succeed in facing difficult problems such as applying machine learning methods.

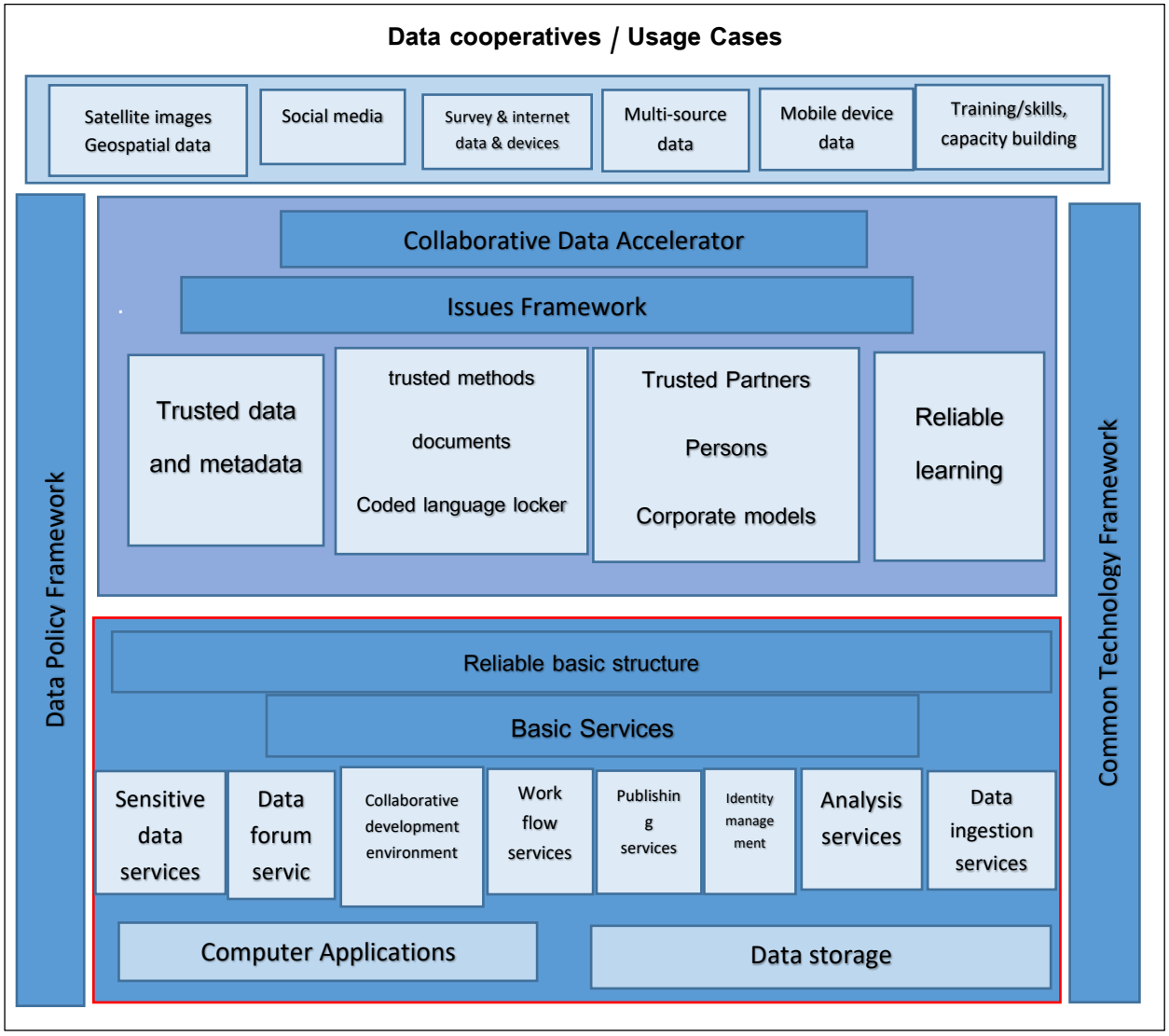
D. **Processing of unstructured data**: Big data tools can process not only data in the form of traditional tables but also unstructured data. Some big data tools (in particular: NoSQL - document databases) can index the content of text files and make it easier to categorize them according to their content. This is useful, especially when calculating statistics from sources such as web pages downloaded from the Internet or social media messages.

There is an important issue that must be taken into consideration. The introduction of new technology solutions to statistical devices requires careful consideration to the method of integrating these technologies into the technical environment, where technicians and statisticians are accustomed to using statistical software tools such as SAS or R. from a technical standpoint, all modern statistical software gives access to big data tools, at least Hadoop, and provides user-friendly interfaces for it. In this sense, big data tools can be used in a similar way to an RDBMS: The big data is handled by big data tools and the statistical program is provided with the refined results which can be considered as inputs for further analysis.

3. Comparing Data Sources

The figure below shows the classification of big data tools according to the data sources considered in the study of products and methodologies. For each data source, Figure 6 shows the recommended tools.

Figure (6): A General Framework for Statistical Data Sources and Structures:



The figure also indicates the specialization of each technology with respect to a particular type of data, structured (vertically with a fixed number of attributes and a spread of numerical values), openly structured (vertically with a variable number of attributes and/or prevalence of textual content) or completely unstructured (raw text or binary files). Diagram and key value databases are not listed in the context of statistical production due to their marginal importance.

4. **Global Team on Big Data in Official Statistics (GWG on BD)**

The United Nations Statistical Commission, at its forty-fifth session in 2014, agreed to establish a global team for the use of big data in official statistics to investigate the benefits and challenges of using big data in the production of official statistics. In this context, the group and the statistical community recognized the need to work on important and necessary issues for the correct and optimal exploitation of big data. On top of these issues is the issue of access to agreed methodologies and legislation related to access to data and technical infrastructure.

The group consists of 9 international organizations:

- ✓ The World Bank
- ✓ UN Global Pulse
- ✓ Department of Statistics of the United Nations
- ✓ United Nations Economic Commission for Europe
- ✓ Economic and Social Commission for Asia and the Pacific
- ✓ Organization for Economic Cooperation and Development
- ✓ International Telecommunication Union
- ✓ Statistical Center for the Cooperation Council for the Arab Gulf States
- ✓ European Statistics

In addition to many countries, where the group was divided into 7 teams as follows:

- Access to data and partnerships.
- Big Data and the Sustainable Development Goals.
- Mobile data.
- Satellite data and geospatial data.
- Data of social networks.
- Training, skills and capacity building.
- Global Platform Committee for Data Services.

5. **Works of the United Nations Economic Commission for Europe:**

- ✓ The United Nations Economic Commission for Europe (UNECE) has worked on a big data project as an initiative of the high-level team on updating the official statistics' systems. The United Nations Economic Commission for Europe (UNECE) project on Big Data aims at the possibility of using big data in official statistics and necessary technologies to take advantage of those sources, in addition to the applicability of current statistical standards and methods to big data.
- ✓ This group adopts a hands-on approach that enables participants to test big data sources and tools on the common collaborative platform called Sandbox. Sandbox can be defined as a technical environment that enables statistical offices in Europe in particular and the world at large to conduct experimental

studies on big data. It is hosted by the Central Statistical Office of Ireland and the Irish Center for Advanced Computing (ICHEC).

The UNECE Data Project consists of five task forces as follows:

- Partnerships Team.
- Privacy Team.
- Quality team.
- Sandbox team.
- Skills and training team.

Conclusion

At the end of the study, it can be said that big data is defined as: “A wide and very complex set of data has its own characteristics (such as size, speed, diversity, variance, validity of data), it is difficult to process it using simple technology to benefit from it. The challenges that accompany this type of data lie in processing, providing, analyzing, storing, searching, sharing, transferring, photographing and updating it, in addition to preserving the privacy associated with it” (Al-Bar, 20147).

The introduction of its approval for use in national statistics has led to a significant importance elicited from the importance of the huge amount of data produced by the digital revolution, which must be controlled in the way it is managed and used in the best possible way by working on important and necessary issues for the correct and optimal exploitation of big data. On top of these issues is the issue of access to agreed methodologies and legislation related to access to data and technical infrastructure.

The scope of use of big data can be categorized into three different types:
Produce new statistical products, supplement existing statistical products, and eventually replace existing statistical products.

In order to achieve this, the researcher recommends the following:

Recommendations:

Based on the results of the study and the findings of the researcher, he recommends the following:

- Building capacities and specialized cadres in the various fields of information technology; to help analyze big data and produce official statistics from it.
- Raising awareness of the importance of big data among workers in the official statistical sector, including leaders and employees in public and private institutions.
- Working on building a national strategy for the development of official statistics using non-traditional data sources, including big data.
- Investing in the data available with various governmental and private agencies and social media through pilot projects to produce official statistics.
- The need for having a national framework for governance and ensuring the security and protection of data privacy for statistical items.
- Focusing on benefiting from data analysis software, determining the needs of the statistical work environment, and matching the type of data with the areas of statistical work that it needs.
- Encouraging the private sector to invest in the development of big data technologies.

Approved References:

First: Arabic References:

1. Ebtissam Hamlawy. (2018). The Reality and Prospects of Big Data in the Arab World: A Case Study of the Kingdom of Saudi Arabia. International Journal of Economic Studies, (3), 122-137.
2. Al-Aklabi, p. (2017). Transforming big data into added value. Journal of King Fahd National Library, 23(2).
3. Al-Bar, p. (20147). Big data and its applications: http://www.kau.edu.sa/GetFile.aspx?id
4. Sabrina Maknani, and Shubila Moghaddam. (2019). The role of big data in supporting sustainable development in the Arab countries. Journal of Information and Technology Studies, 1 - 14.
5. Al-Tayeb, Z., & Al-Rubai, S. (2018). The new roles of the information specialist to deal with big data. Journal of Information Studies & Technology, 16, 1-15.
6. The Organization for Cooperation and Development. (s.d.). OCDE: http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm
7. Information and communication technology. (Ministry 2014). Big data: balancing benefits and risks. Qatar.: Ministry of Information and Communication Technology.

Second: English References:

8. Canopy, L. (2015). Big data for marketing: When it is big data the right choice. Récupéré sur Big data for marketing: When it is big data the right choice: file///c/:/Users/user/Downloads/canopyfile-whitepaper_big-ata-formarketing.pdf.
9. Meguneni, S., & Mokedem, C. (2018). Role of Big Data in Supporting Si=ustainable Developpement in Arab Countries. Recovered on Journal of Information and Technology: https://doi.org/10.5339/jist.2019.
10. Mercier, M. (2019). Contribution to High Performance Computing and Big Data Infrastructure Convergence. Grenoble: Université Grenoble Alpes.
11. Sanders, J. (2016). Defining Terms: Data, Information and Knowledge.
12. Ward, B. (2004). “How Linux Works: What Every Superuser Should Know?”.
13. Weiss, A., A. (2018). Big Data Shocks: An Introduction to Big Data for Librarians and Information Professionals. Chicago: American Library Association.