STATISTICS IN TRANSITION new series and *STATISTICS OF UKRAINE, February 2023*
Joint Special Issue: A New Role for Statistics
Vol. 24, No. 1 pp. 27–41, DOI 10.59170/stattrans-2023-003
Received – 01.07.2022; accepted – 12.10.2022

Using Big Data by Ukrainian official statistics when martial law applies: problems and solutions

Oleksandr H. Osaulenko[1], Olena Horobets[2]

ABSTRACT

The article is focused on issues of the secure operation of official statistics in Ukraine during the application of martial law. It was found that the gaps in conventional sources of statistical data caused by the war needed to be filled with data from alternative sources, including Big Data. The level of digitalisation in Ukraine as the basis for using Big Data was analysed by the proposed indices of internetisation, social progress and digital transformation. Thanks to our research, several problems (methodological, legal, financial, and managerial) were identified as vital for statistical offices on their way to the implementation of Big Data in statistical processes. Our proposals concern tools for Big Data processing, such as Data Hypercube as a way for presenting Big Data for their visualisation, applications of Web scraping in estimating the consumer prices index, analyses of labour and real estate markets, and the applications of specialised software for the collection, processing and analysis of Big Data sets

Key words: official statistics, statistics during war, Big Data, digitalisation.

1. Introduction

In connection with the large-scale invasion of the territory of sovereign Ukraine by the Russian Federation, the President of Ukraine signed the Decree "On the Imposition of Martial Law" and a series of other legal acts, including the Law of Ukraine "On the Protection of Interests of the Entities Submitting Reports or Other Documents in the Period of Martial Law or the State of War", which specifies that the submission of any category of reports in paper or electronic form, including statistical and financial reports, shall be suspended for the period of martial law and three months following its abolition.

[1] National Academy of Statistics, Accounting and Audit, Kyiv, Ukraine. E-mail: O.Osaulenko@nasoa.edu.ua. ORCID: https://orcid.org/0000-0002-7100-7176.
[2] National Academy of Statistics, Accounting and Audit, Kyiv, Ukraine. E-mail: babutska@ukr.net. ORCID: https://orcid.org/0000-0003-1762-2140.

Apart of from irreversible losses for Ukraine and each of its residents, the war has hit its public administration system severely, including official statistical bodies, which had to suspend the collection of primary statistical data and publishing of statistical information.

In fact, the official statistics of Ukraine now is devoid of all the main sources for statistical information: primary data obtained through statistical observations and surveys (coming from business entities, physical persons, households, as well as government organizations, public entities, etc.), secondary data from administrative sources, and information that used to be collected by specialized government agencies.

In view of the above, it can by suggested that the operation of the Ukrainian system for primary statistical data collection has actually been stopped. Given that the availability of reliable statistical data constitutes an integral component of the efficient work and progress in all the walks of life, public administration included, with the official statistical information being part of the global information space, the official statistics of Ukraine urgently needs the involvement of alternative data sources and applications of innovative methods for their processing, in order to fill the gaps in the conventional data environment and preserve the national statistical system.

2. Methodology

When writing this article, the authors used methods of scientific theory and statistics, such as theoretical generalization, systematization and index analysis (to assess the digital transformation process).

3. Big Data: a promising source for statistical data in time of war

Although Big Data have gained increasing popularity in the recent years, this notion had long been in use before it became conventional. This term was used in the context of visualizing Big Data sets as early as 1997 (Cox and Ellsworth, 1997). It was applied in the context of data analysis in 1998 (Weiss and Indurkhya, 1998), and in the context of statistics in 2003 (Diebold, 2003).

The Oxford English Dictionary (2022) defines Big Data as “sets of information that are too large or too complex to handle, analyse or use with standard methods”; the Gartner Dictionary (2022) – as “high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation”. It follows that Big Data refer to big sets of data on diverse social and economic phenomena and processes, produced on a continuous basis.

Big Data are characterized by specific components, which are six in number as distinguished by Microsoft Company. These components are given by C. Wu, R. Buyya, K. Ramamohanarao (2016):

1. Volume stands for scale of data.
2. Velocity denotes the analysis of streaming data.
3. Variety indicates different forms of data.
4. Veracity focuses on trustworthiness of data sources.
5. Variability refers to the complexity of data sets. In comparison with "Variety" (or different data format), it means the number of variables in data sets.
6. Visibility emphasizes that you need to have a full picture of data in order to make informative decision.

As argued by Bentein A. (2021), "the amount of digital data globally is doubling approximately every three years. The 74 zettabytes we have today stand for 79 billion terabytes, which is more than 10 terabytes for each living human on earth". Wiltshire D. & Alvanides S. (2022) emphasize that "Big data holds great potential for research and for society, large volumes of varied data can be produced and made available to researchers much faster compared to 'traditional' data… Big data is generated in higher frequencies than other forms of data, such as from social surveys and national censuses that can take months even years to be made available to researchers". It follows from the above that Big Data can provide a potential source for more relevant and timely statistical information compared with its conventional sources.

The global statistical community has realized the great significance of Big Data. The crucial step towards it was made by the Conference of European Statistical System on Big Data Problems (ESS Big Data event, 2014) in Rome, where the issue of statistical uses of Big Data was defined as a core one for the official statistics.

Later on, the European Commission clearly articulated its vision of Big Data sets as an important resource for economic growth and social progress on the whole. In particular, the European Strategy for data (2020) states that "data is at the centre of this transformation and more is to come. In a society where individuals will generate ever-increasing amounts of data, the way in which the data are collected and used must place the interests of the individual first, in accordance with European values, fundamental rights and rules. At the same time, the increasing volume of non-personal industrial data and public data in Europe, combined with technological change in how the data is stored and processed, will constitute a potential source of growth and innovation that should be tapped".

On February 23, 2022, the European Commission approved the Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (2022), with the aim of ensuring fairness in the allocation of value from data among actors in the data economy and to foster access to and use of data.

Big Data types are classified according to a definition that is mostly based on data sources, (United Nations, 2013), as follows: "Data and information sources arising from the administration of a programme, be it governmental or not, e.g., electronic medical records, hospital visits, insurance records, bank records and food banks; Commercial or transactional sources arising from the transaction between two entities, e.g., credit card transactions and online transactions (including from mobile devices); Sensor network sources, e.g., camera data, satellite imaging, road sensors and climate sensors, such as those pertaining to Remote Sensing data sources; Tracking device sources, e.g., tracking data from mobile telephones and the Global Positioning System (GPS); Behavioural data sources, e.g., online searches (about a product, a service, or any other type of information) and online page views; Opinion data sources, e.g., comments on social media; Geographic information system (GIS) data and information of various sources and types".

Following the adoption of the Sustainable Development Goals (SDGs), National Statistics Offices worldwide are requested to undertake a data revolution, which implies a titanic challenge to their mandate, namely to populate the SDGs matrices in all their economic, social and environmental dimensions (United Nations, 2021).

In spite of the abovementioned, a standard algorithm for collection, grouping, refining and analysis of Big Data still does not exist, a fact entailing distortions of produced results and formulating erroneous conclusions when processing these data sets. Apart from due consideration to the scopes, velocity, diversity, authenticity and variability of Big Data flow, its complex ecosystem has also to be born in mind. It follows that another characteristic of Big Data should be added, the hierarchical complexity caused by the full reliance on ICTs, whose advancement stimulates the occurrence and diversification of unique data in the continuous process. It involves the creation of a single set of primary (raw) data that can be roughly divided into three chains, with each one containing several interrelated groups, thus creating the data ecosystem.

The first chain contains the groups of data from general administrative data to data reflecting individual or public opinions or sentiments. The second chain, being more granulated and personalized, contains, inter alia, confidential data that have synergetic effects for each other. The third chain in the ecosystem, operating as a data filter, contains the already accurate and harmonized data formed in the four global groups: Shallow data, Deep data, Micro-data, and Nano-Data. It should be added that each type of the data has the so-called Dark data (Gartner, 2022a) of its own, which remain after mining. Dark data are not meant for repeated use because of high cost for their handling, which puts in question the reasonability of these data storage.

In this context, Henriques A. (2022) emphasizes that analysts are becoming more and more concerned with the occurrence of structured, non-structured and semi-structured data of a new type in the ecosystem, which are the data of experience, accumulating the data on education, life activities, behaviour or performance. Thus, the data on life activities reflect, for the most part, the specifics of routine life of a human and his/her daily interactions; the data on education contain evidence on formal or informal education, gained skills or competencies, etc.; the data on behaviour inform on the efficiency of a human in the society, helping reveal his/her potentials or threats; the data on performance give evidence of success on a job place (with sources of these data being ratings, results of interviews, etc.). Although analysis of data on experience still remains at the primary phase, their exclusive importance has already been understood: laying grounds for organization of analytical effort covering all the experiences accumulated by the humanity, for formulating logical conclusions and using results for various practical purposes, including the elimination of the warfare impact on the official statistics and the society on the whole.

When elaborating methodologies for the implementation of Big Data as alternative sources for the official statistics, some problems pertaining to this environment should be put in focus.

Cai et al. (2019) emphasize the problems related to data collection, storage and analysis queries. Kenglung H. (2022) prioritizes the access to reliable information in time of Big Data collection and analysis.

Making Big Data applicable for statistical problems requires the elaboration of a radically new policy on data management and protection, relations with respondents and the process of professional training of statistical staff.

With respect to the methodology, the problem involved in Big Data use in statistics is how to ensure their representativeness. Given the lack of a formalized general population as such, difficulties in defining a targeted population on the whole and sampling can occur.

It is known that conventional statistical observations are based on censuses and registers allowing to define all the types of populations required for conducting a statistical observation. However, with Big Data formed spontaneously, their random, unstructured and dynamic nature complicates the statistical process that is supposed as stable for the official statistics concept. Besides, the existing statistical methods are designed for a consistent, deep and long-lasting analysis of data in small sized samples, which is admitted to be a strong impediment for the statistical production process.

Another methodological problem in using Big Data for statistical purposes is how to ensure the relevance of technical support for statistical studies with respect to data quality measurement, the overall limitation of applications of data from external sources, the complexity of integrating information from different external origins in statistical databases to have a reliable end product.

It should be added that there also exist legal problems related with the regulation of access and use of Big Data, first and foremost with protection of personal data of residents and individual data of business enterprises, financial problems of optimizing the ratio of costs and gained advantages, administrative problems in policy-making with respect to the principles of data management and protection.

In the context of this article's theme, it should be noted that the warfare in Ukraine features the intensive involvement of information and communication technologies, especially ones based on artificial intellect, operated by the use of effectively designed and customized algorithms for Big Data processing.

In view of the above, it should be emphasized that smart technologies based on artificial intellect are popular in Ukraine, whose data are incorporated in one of the above mentioned global groups of Big Data ecosystem. Smart technologies are largely used by Ukrainians at household level, being installed in houses, cars or house territories. These technologies[3] have been applied by the Ukrainian Ministry for Digital Information to collect infinite numbers of video and photo evidence of violations of laws and customs of war, fixed in the agreements of Geneva Convention (Geneva, 1949).

Another essential source for Big Data is social networks. Their popularity, together with technologies based on artificial intellect, have allowed to find and identify servicemen of the Russian armed forces engaged in brutal massacres of civilian population in Ukraine in the towns of Bucha, Hostomel, Irpin in the Kyiv region of Ukraine (Ministry of Digital Transformation of Ukraine, 2022).

This successful experience in Big Data applications in the conditions of warfare gives evidence of the value and timeliness of this data type, in spite of the persisting "problematic" issues pertaining to their implementation, including in official statistics.

4. Digital framework for using Big Data in Ukraine

The implementation of Big Data in any kind of life activity calls for a certain level of technological readiness of a country as a whole and its sectors, with the adequate level of digitalization in the first place. Digitization is a current megatrend, meaning that digital technologies are integrated into our everyday life. The use of digital technologies enables the connection of different services and automation of many processes.

According to the OECD report (2017), the modern society has been a witness of the gradual expansion of digitalization that has impact on all the activities, it is on the way to the digital economy. This process began nearly 50 years ago, with the recent years marked by an incredibly higher pace at which digitalization penetrates the life.

[3] Titles and components of used technologies based on artificial intellect are not indicated by the authors for security considerations.

When reference is made to future scenarios of the social development, it is emphasized in the report that nowadays the world has plunged in the digital megalomania. First and foremost, it is caused by the rapid technological advancement in information technology sector and the rapid diffusion of Internet, entailing the increasing scopes of information, a kind of non-stop data. Mayer-Schönberger and Cukier (2013) labelled this phenomenon as datafication.

It should be noted in view of the above that Ukraine could create a positive self-image in the digital space. It was confirmed by estimations of Internetization Index proposed by the authors (Table 1) and Digital Transformation Index (Table 2).

Internetization Index is based on nine indicators: number of Internet subscribers, contention ratio, scope of provided Internet services, availability of the following ICT devices in households: personal PC, notebooks, Ipads, mobile phones, ratio of technological development, ratio of the availability of advanced technologies in Ukraine.

Because the data for these indicators are stimulators, their normalization was made by the maximal value over the period by the formula:

$$z_{ij} = \frac{x_{ij}}{x_{i,max}},$$

where x_{ij} is the value of *i* indicator in *j* year, $x_{i,max}$ is the maximal value of *i* indicator in the period under study.

Internetization Index was estimated by the formula:

$$I_j = \frac{\sum_i z_{ij}}{n},$$

where I_j is the value of Internetization Index in *j* year, n is the number of indicators determining Internetization Index, $n = 9$.

Table 1. Normalized indicators and Internetization Index

Year	z1	z2	z3	z4	z5	z6	z7	z8	z9	$\sum_i z_{ij}$	Internetization index, %
2016	0.642	0.663	0.725	1.000	0.771	0.789	0.990	0.850	0.956	7.387	82.1
2017	0.881	0.788	0.815	1.000	0.771	0.789	0.990	0.900	0.956	7.890	87.7
2018	0.907	0.894	0.893	0.974	1.000	1.000	1.000	0.950	0.978	8.595	95.5
2019	1.000	1.000	1.000	0.974	1.000	1.000	1.000	1.000	1.000	8.974	99.7

Data in Table 1 show that in the pre-war Ukraine there were favourable conditions for the penetration of digitalization.

But a good level of Internet penetration is not sufficient for the assessment of the society's readiness for global digitalization. Bearing this in mind, the authors proposed Social Progress Index to be used as a balance indicator, which is based on a set of indicators measuring the social development of a country in an integrated way. SDI used in this study consists of the indicators characterizing three areas of the social development: basic human needs, foundations of well-being, and opportunities (Global Index).

Basically, Social Progress Index can cover data on 60 indicators, including: access to information and communications, freedom of choice and personal freedom, access to basic education, access to higher education, personal security, tolerance, nutrition, basic medical service, quality of environment, etc.

A comprehensive description of the society's readiness for digitalization processes was made by the authors with use of Digital Transformation Index that was estimated as the arithmetic mean of the two sub-indices: Internetization Index and Social Progress Index.

Table 2. Sub-indices and Digital Transformation Index, %

Year	Internetization Index	Social Progress Index	Digital Transformation Index
2016	82.1	66.4	74.3
2017	87.7	68.4	78.1
2018	95.5	69.1	82.3
2019	99.7	66.9	83.3

Data in Table 2 show that the society's readiness for digital transformation in Ukraine was increasing year by year in the pre-war period, giving evidence that the country's ability to face the challenges of the modern digital world.

Thus, the EasyWay service (see: https://www.eway.in.ua) informing on all the routes and stops of public transport in 60 cities of Ukraine, as well as Poland, Moldova, Bulgaria. Uzbekistan, Serbia, Croatia, and Kazakhstan, was developed in Ukraine as early as in 2011. The main feature of this service is online monitoring of the movement of buses, trolleybuses, trams and urban trains via GPS.

It was in 2019 that the Ministry of Digital Transformation of Ukraine presented a mobile application and web portal under "Diia" brand. This application allows one to keep driver identification, internal and external passports, vaccination certificates (Covid-19) and other documents in one's smartphone, and to transmit their copies when receiving bank or post services, in hotel check-in or in other circumstances of daily life. Also, this application opens free access to some categories of public services: one can receive the comprehensive service in time of child birth (eMaliatko (eBaby)), register business online, pay taxes and submit statements, sign any category of documents, change the place of registration, etc.

It is expected that 100 percent of public services will be provided via "Diia" (see: https://diia.gov.ua/) until 2024. At the end of 2021, the application and portal was used by more than 12 million persons. The portal has provided access to 72 services, and the application – to 9 services and 15 digital documents. An important feature of the application is that it does not collect and store individual data of Ukrainians, being designed only for displaying the data filled in the official registers.

The brand has incorporated the newly created and already implemented public project "Diia City": it is a fiscal and regulatory space for IT companies, designed for providing tax privileges and opportunities to apply advanced tools for attracting venture investments or protecting intellectual property. Earlier this year "Diia City" was included in the short list of Emerging Europe Awards 2022 in "Partnership" category, nomination "Modern and Future-Proof Policy".

Also, Ukraine has successfully implemented the program for digitalization of health protection through designing and introducing a two-component system by which a user can interact with the central database via the online medical information system eHealth (see: https://ehealth.gov.ua/), which enables to computerize the operation of business entities in the heath protection sector, to create and review information, and to exchange information in electronic form.

Mobile operators in Ukraine have already launched mass-scale applications of Big Data for providing assistance to business: creating a client portrait, look-alike audience, target, trigger emails, geoanalitics, and financing scoring.

Four days after the beginning of the war, following the request of the Minister of digital transformation, Elon Musk opened access to high-speed Internet for Ukraine: his Starlink is providing assistance for the militaries and supporting the operation of critical infrastructures and services in medical care, finance and energy supply.

The abovementioned allows to admit that the expansion of digital processes across the Ukrainian territory lay a firm ground for using Big Data that have been gradually penetrating various walks of life.

5. Proposals

The main effort of Ukrainian statisticians now needs to focus not only on building the Big Data capacity or exchange of knowledge at international scales, but, first and foremost, on developing partnerships with the private sector for practical applications of gained knowledge and skills with the final end of involving and processing these new category of data. In the conditions of war, Big Data should become a supplement to the conventional statistical data or their alternative with an even higher level of detailing.

There is a large number of proposals on using Big Data in the official statistics. The process of implementing Bid Data in the official statistics must necessarily conform to the Fundamental Principles of Official Statistics (2014). A useful approach will be the reliance on the principles for Big Data analysis (Chen and Zhang, 2014) and 12 rules of Codd (1985), which have already been classical in building database management systems.

In view of the abovementioned and considering the necessity of using Big Data by the official statistics, we propose the following principles of Big Data handling:

1. Big Data collection procedure must guarantee their security and confidentiality, and must be backed by tests for reliability of quality of data.
2. Big Data collection and analysis, publishing and dissemination of the conclusions based on the results of their processing must have intellectual support.
3. The norms and rules pertaining to Big Data collection, processing and analysis, publishing and dissemination of the conclusions based on the results of their analysis must be agreed and publicized in the format of quadripartite group "government-science-business-statistics".
4. The guaranteed free access to official and private Big Data repositories for members "government-science-business-statistics" group.
5. Statistical departments must elaborate methodologies and recommendations on Big Data collection, processing and analysis considering the specific features of each Big Data group.
6. The algorithms for main processes of Big Data collection, processing and analysis must be made consistent and open to broader public.

Once the above principles are followed, it will allow for further elaboration and unification of methods on the way from "unordinary proposals on using Big Data" towards "adequate management decisions on the implementation of Big Data in various walks of life".

A greatly important factor in decision-making pertaining to statistical uses of Big Data is the choice of proper tools for their handling, first of all in IT technologies with applications of mathematical modelling methods.

A method of Big Data presentation that seems appropriate for statistics is Data Hypercube (Marchand, P., Brisebois et al., 2003), a unique organization of data in the form of a multidimensional set, the so-called cube of data.

The technique for this organization is a component of On-Line Analytical Processing (Codd et al., 1993), a technology for analytical data processing by the use of methods for collection, storage and analysis of multidimensional data for the support of decision-making processes. The cube structure allows for making a quick and multifaceted analysis from various perspectives, which can be fit for Big Data presentation for their further visualization.

An important step in the implementation of Big Data in the conventional statistical processes is the choice of appropriate computerized tools, applications and software required for these data handling.

A quite effective tool for extracting the necessary data is Web scraping. According to Dogucu M. & Çetinkaya-Rundel M. (2020), "Web scraping is the process of extracting data off the web programmatically and transforming it into a structured dataset. Web scraping allows for larger amounts of data to be collected in a shorter span of time and in an automated fashion that minimizes errors". Web scraping can be efficiently used in estimating the consumer price index (using websites of retail trade stores), analysis of the real estate market, etc.

However, this tool has weak sides. Summa D. et al. (2019) mention some problems that may be faced by statisticians in time of web scraping. They argue that online price data (web scraped) from the e-commerce platforms may not cover the full list of goods or services that the NSOs rely on in the compilation of the CPI. Technical issues include frequent changes in the website structure, the need to update crawlers or develop separate crawlers for different websites, the possibility of automatic blockage of high frequency web scraping, which calls for collaboration and partnership with web site owners. Another important challenge is the quality of data.

In spite of this, the experience of Italian National Institute of Statistics (Polidoro et al., 2015) in using Web scraping for the elaboration of new strategies for integrating data from conventional surveys, administrative bodies and innovative sources like Big Data, give evidence of positive results in reducing the statistical burden on respondents, allowing to create relevant databases and new formats of statistical registers.

It should be noted that the rapid expansion of digitalization processes stimulated many software developments for data analysis, which can be used for statistical purposes: Actian (a tool allowing to store raw data and their preparation for further analyses), Ambari (a tool for cluster management), Avro (a system of data serialization), Apache Kafka (a platform for data processing in the real-time format), Hive (an infrastructure for data storage, enabling for their aggregation), etc.

One of the high performing tools for analysis of Big Data with the open code is a collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of Big Data using the MapReduce programming model – Apache Hadoop. A notable advantage of Hadoop is easy storage and dissemination of large data sets on servers that can be operated in parallel, with data management by the use of clusters, thus enabling a method for storage based on shared file systems.

6. Conclusions

The results of this research could demonstrate that the needs of the official statistics in primary data can be effectively met in time of the martial law through using alternative sources, Big Data in the first place. These data should be introduced in parallel or in mix with conventional data sources, to fill the gaps in conventional data due to the war.

Nowadays, the important Internet sources of Big Data in Ukraine are applications designed by the government sector: "Diia", eHealth, the platform of public procurement (Prozorro), the system of road motion sensors, consumption meters. Also, Ukraine has an extensive network of private digital services: online banking services (e.g. Monobank, the first neobank in Ukraine), a logistic application for express delivery ("Nova poshta"), market places (Prom, Rozetka, Bigl), etc. More "familiar" sources of Big Data (mobile phones, social networks, Google analytics, etc.) have to be considered, too.

Acknowledgement

The authors express sincere gratitude to the defenders of Ukraine for the opportunity to write this article in conditions that are safe for life.

References

Bentein, A., (2021). Data is the New Gold. Retrieved from: https://www.qad.com/blog/2021/04/data-is-the-new-gold#:~:text=Are%20the%20phrases%20%E2%80%9Cdata%20is,spend%20more%20money%20on%20technology%3F

Cai, L., Qi, Y., Wei, W., Wu, J., Li, J., (2019). mrMoulder: A recommendation-based adaptive parameter tuning approach for big data processing platform Future Gener. Comput. Syst., 93, pp. 570-582.

Codd, E. F., (1985). Is Your DBMS Really Relational? ComputerWorld.

Codd, E. F., Codd, S. B. and Salley, C. T., (1993). Providing OLAP to User-Analysts: An IT Mandate. Retrieved from: https://web.archive.org/web/20170808214004/https://www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf

Cox, M., Ellsworth, D., (1997). Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th conference on Visualization'97, pp. 235–247.

Chen, C. L. Philip and Zhang, C.-Ya, (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Science, 275, pp. 314–347.

Diebold, F. X., (2003). Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting. (Discussion of Reichlin and Watson papers), in M. Dewatripont, L.P. Hansen and S.Turnovsky (Eds.), Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society. Cambridge: Cambridge University Press, pp. 115–122.

Dogucu, M., Çetinkaya-Rundel, M., (2021). Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities. Journal of Statistics and Data Science Education, Vol. 29, pp. 112–122.

ESP Rome, (2014). Collaboration in Research and Methodology for Official Statistics (CROS). Retrieved from: https://ec.europa.eu/eurostat/cros/content/esp-rome-2014_en.

European Commission, (2022). Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data. Retrieved from: https://digital-strategy.ec.europa.eu/en/library/data-act-proposal-regulation-harmonised-rules-fair-access-and-use-data.

European Commission, A European Strategy for data. Retrieved from: https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

European Commission. (2020). Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the Committee of the regions a European strategy for data. COM/2020/66 final. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066.

Gartner Glossary, (2022). Big Data. Retrieved from: https://www.gartner.com/it-glossary/big-data/.

Gartner Glossary, (2022a). Dark Data. Retrieved from: https://www.gartner.com/it-glossary/dark-datahttps://www.gartner.com/it-glossary/dark-data.

Geneva, (12 August 1949). Convention (IV) relative to the Protection of Civilian Persons in Time of War. Retrieved from: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Treaty.xsp?documentId=AE2D398352C5B028C12563CD002D6B5C&action=openDocument.

Global Index: Overview. Social Progress Imperative. Retrieved from: https://www.socialprogress.org/index/global.

Henriques, A., (2022). Mind the gaps: how experience data can help fight climate change. World Economic Forum. Retrieved from: https://www.weforum.org/agenda/2022/02/climate-change-experience-data/.

Kenglung H., (2022). Big data analysis and optimization and platform components. Journal of King Saud University – Science. Vol. 34. Issue 4. Retrieved from: https://www.sciencedirect.com/science/article/pii/S1018364722001264#b0015.

Marchand, P., Brisebois A., Bédard Y. and Edwards, G., (2003). Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis. Workshop ISPRS, Québec, Canada.

Mayer-Schönberger, V., Cukier, K., (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt. Retrieved from: https://psycnet.apa.org/record/2013-17650-000/.

Ministry of Digital Transformation of Ukraine, (2022). Mykhailo Fedorov explains about recognizing the faces of the Russian occupiers. [Mykhaylo Fedorov poyasnyuye pro rozpiznavannya oblych rosiys′kykh okupantiv]. Retrieved from: https://www.kmu.gov.ua/news/mihajlo-fedorov-pro-rozpiznavannya-oblich-rosijskih-okupantiv [in Ukrainian].

OECD, (2017). Going Digital: Making the Transformation Work for Growth and Well-Being: Meeteng of the OECD Council at Ministerial Level. Retrieved from: https://www.oecd.org/mcm/documents/C-MIN-2017-4%20EN.pdf.

Oxford Learner's Dictionaries, (2022). Big Data. Retrieved from: https://www.oxfordlearnersdictionaries.com/us/definition/english/big-data.

Polidoro, F., Giannini, R., Conte R. L., Mosca, S. and Rossetti, F., (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. Statistical Journal of the IAOS, 31(2), pp. 165-176.

Summa, D., Bianchi, G., Consalvi, M., Gentili, B., Pancella, F. and Scalfati, F., (2019). Using Big Data for Official Statistics: Web Scraping as a Data Source for Statistical Business Registers (SBRs). NTTS2019. Retrieved from: https://coms.events/ntts2019/data/abstracts/en/abstract_0007.html.

United Nations, (2013). Big Data and modernization of statistical systems. United Nations Economic and Social Council. Report of the Secretary-General. Forty-fifth session of the UN Statistical Commission, NY, 4-7 March 2014, doc. E/CN.3/2014/11, December, New York, pp. 1–16.

United Nations. Economic and Social Commission for Western Asia, (2021). Use of Big Data in Compilation of SDG Indicators in the Arab Region, Challenges and Opportunities. Retrieved from: https://www.unescwa.org/sites/default/files/pubs/pdf/big-data-compilation-sdg-indicators-arab-region-challenges-opportunities-english.pdf.

United Nations Statistics Devision, (2014). Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014). Retrieved from: https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

Weiss, S. M., Indurkhya, N., (1998). Predictive data mining: a practical guide. Morgan Kaufmann Publishers, 1997, pp. 44–48.

Wiltshire, D., Alvanides, S., (2022). Ensuring the ethical use of Big Data: lessons from secure data access, Heliyon, Vol. 1. Issue, 2. pp. 1–6.

Wu, C., Buyya R., Ramamohanarao, K., (2016). Chapter 1. Big Data Analytics = Machine Learning + Cloud Computing. Big Data Principles and Paradigms, Elsevier, pp. 3–38.