УДК: 311.17:004.6
JEL Classification: C4; C15;
doi: 10.31767/nasoa.3-4-2021.03

O. OSAULENKO,
Dsc (Public Administration), Professor,
Correspondent Member of the National Academy
of Sciences of Ukraine, Rector,
e-mail: o.osaulenko@nasoa.edu.ua,
ORCID ID: 0000-0002-7100-7176,
ResearcherID: F-3856-2018;
O. HOROBETS,
PhD (Economics);
National Academy of Statistics, Accounting and Audit;
e-mail:babutska@ukr.net,
ORCID ID: 0000-0003-1762-2140,
ResearcherID: G-7664-2018

Social Media Data in the Big Data Environment

The article contains results of a study of social media data (SMD) which, being distinct from conventional data by their origin, require special methods for collection, processing and analysis. As shown by a literature review, in spite of great many research publications devoted to social media research and big data analysis, the SMD potential as a big data component still remains inadequately explored.

Two approaches to research and analysis of SMD were highlighted in course of the study, in which SMD are addressed as an object of Internet statistics and an object of big data. When SMD are explored as an object of Internet statistics, collection of anonymized data is performed using the services that have network protocols for collection and analysis of data on social media customers using statistical methods. When SMD are explored as an object of big data, the collection is performed mostly by artificial intellect, whereas the storage and processing is operated by databases designed for large scopes of data and software with statistical data processing applications.

The social media most popular with users in 2020 were identified in the study. Statistical indicators for assessment of users' feedback, available now for statistical assessments of social media communities, are given. The study revealed several problems which solutions would require, apart from a multifaceted and complex approach to collection and processing, highly competent teams of specialists in various subject fields, including experts in computations, experts in machine learning and statisticians.

Keywords: *social media data, big data, social media, Internet statistics, public sentiments.*

О. Г. ОСАУЛЕНКО,
доктор наук з державного управління, професор,
член-кореспондент НАН України, ректор;
О. О. ГОРОБЕЦЬ,
кандидат економічних наук;
Національна академія статистики, обліку та аудиту

Дані соціальних медіа у середовищі великих даних

У статті розглядаються дані соціальних медіа (ДСМ), які за походженням відрізняються від традиційних даних і тому потребують особливих методів збирання, оброблення та аналізу. Визначено, що незважаючи на велику кількість наукових публікацій, які присвячені питанням вивчення соціальних медіа і аналітики великих даних, потенціал ДСМ як складової великих даних усе ще залишається недосконало дослідженим.

У ході дослідження виокремлено два підходи до вивчення та аналізу ДСМ, згідно з якими ДСМ розглядаються як об'єкт Інтернет-статистики і як об'єкт великих даних. При дослідженні ДСМ як об'єкта Інтернет-статистики збирання знеособлених даних здійснюється за допомогою сервісів, які мають мережеві протоколи для збирання і аналізу даних про відвідувачів соціальних медіа з використанням статистичних методів. Якщо ДСМ досліджуються як об'єкт великих даних, збирання здійснюється переважно за допомогою штучного інтелекту, а для зберігання та оброблення використовуються бази даних, розраховані на великі обсяги даних, і програми статистичного оброблення даних.

В рамках дослідження визначено найпопулярніші соціальні медіа серед користувачів у 2020 р. Наведено статистичні показники для оцінювання зворотного зв'язку з аудиторією, які наразі доступні для статистичного оцінювання спільнот у соціальних медіа. Виявлено низку проблем, які, як і у випадку з великими даними, вимагають окрім багатоаспектного та комплексного підходу до збирання та оброблення ще й висококваліфікованих команд фахівців з різних предметних галузей, зокрема експертів з обчислень, експертів з машинного навчання, статистиків.

Ключові слова: *соціальні медіа, дані соціальних медіа, великі дані, Інтернет-статистика, соціальні настрої.*

Introduction. Social media has become an integral part of the common people's life and core communication platforms. They help one learn news of friends' life, latest world events, to share own impressions, emotions or thoughts, to get advice or to waste free time. A specific feature of social media is that it reflects public sentiments in real time. Besides that, by using Internet improves one's awareness of the society, economics or politics, stimulates the social activity and create an illusion of the direct involvement of Internet users in all the events and the significance of their opinions displayed in their comments or messages.

The quarantine restrictions caused by the COVID-19 pandemic have increased the audience of social media. From July to September 2020 the audience of social media grew by 180 million in relation to the analogous period of 2019 [1]. A vast scope of various data is generated b users each day, thus creating a new type of data in the big data environment.

In spite of the great many scientific publications devoted on social media and big data analysis, the potential of social media data as a big data component still remains inadequately explored.

Literature review. According to J. Schwaiger et al., "the digital transformation, with its ongoing trend towards electronic business, confronts companies with increasingly growing amounts of data which have to be processed, stored and analyzed. Instant access to the 'right' information at the time it is needed is crucial and thus, the use of techniques for the handling of big amounts of unstructured data, in particular, becomes a competitive advantage. In this context, one important field of application is digital marketing, because sophisticated data analysis allows companies to gain deeper insights into customer needs and behavior based on their reviews, complaints as well as posts in online forums or social networks" [2].

M. Sarprasatham, who analyzed the issue of social media in the bid data environment, argued that "information technology has reached its pinnacle, with the era being dominated by two hi-tech driving forces – Big data and Social media" [3].

Big data allow to determine the tendencies of social media and extract the statistics that can be used in taking decisions on future actions. Updated statuses, photos or videos displayed by users in social media contain valuable information about demographic aspects, likes and dislikes. So, big data enable for not only easy grasp into the sentiments of potential markets, but to build competitive strategies at company level.

Recent publications containing social media research tend to be focused on the business organization issues [4], peculiarities of public sentiments in times of election campaigns [5, 6, 7, 8], developments in the public health sector [4] and essential principles of information security in time of the COVID-19 pandemic.

It was in 2008 that A. Mayfield described social media as a cluster of online portals that help their users establish feedbacks with the audience [8] and form public sentiments in a purposeful manner, which can be observed in time of election campaigns.

An earliest classical example of using social media in political circles with communication purposes was U.S. presidential elections in 2008, being the first time when social media (Facebook, MySpace, Twitter, Flickr, Digg, BlackPlanet, LinkedIn, AsianAve, MiGente, Glee etc.) were used for voters' feedback in a way to organize a highly effective "web-blitzkrieg" of B. Obama.

In view of the above, it can be suggested that social media quite often turn into self-PR platforms better known as blogs. O. Zernetska whose argument is worth a mention in this context notes that "a blogosphere is created in each country, and in the planetary scale – thanks to Internet and the opportunity of each Internet user to read a blog in every corner of the world and send own comments (not to mention active bloggers creating political content in their blogs) – it can be said that the global political blogosphere has been created" [10].

McAllister argues that use of Internet, political awareness and political participation are linked with each other. According to this author, this tendency is critical in time of elections, especially in developing countries [11].

An interesting experience is use of social media in Slovakia in time of the political campaign in 2019 [6]. An analytical tool for Facebook pages, Facepager, was used for analysis of the total number of messages, their types and feedback coefficients for each candidate's pages. As a result, the importance of using social media in time of elections could be confirmed by statistical methods.

It should be noted that social media help disseminated false (fake) information. In view of this the World Health Organization (WHO) defined this phenomenon as "infodemic", i. e. the surplus of online and offline information pertaining to the pandemic, including shares of false information that threatens physical and psychic health of people and endangers the effective implementation of the measures for pandemic control. Also, WHO calls for simplification of the population's access to reliable information flow, with emphasizing that COVID-19 is the first pandemic in the history in which technology and social media are being used on a massive scale to keep people safe, informed, productive and connected [12]. In 2020, in time of the World Health Assembly, WHO member states approved the resolution WHA73.1 on COVID-19 response, in which they called the states to provide a reliable content on COVID-19, take measures for counteracting false information disinformation, and use digital technologies for the response framework [13].

Alton M. K. Chew and Dinesh Visva Gunasekeran who explored dissemination of fake information about the pandemic argued that social media had been a defining component of life in the 21st century, monetizing peer-to-peer sharing of information. This led to the formation of powerful platforms leveraging artificial intelligence (AI) to effectively commoditize individual attention [14].

Interesting enough is a study of the public perception of economic measures imposed in response to the COVID-19 pandemic in Poland in March-June 2020 [9]. This study involved an analysis of big data extracted from Facebook and Twitter, namely 109,022 twits and 557,473 messages in Facebook, with exploring two variables for estimating the public perception of economic support proposed by the government: number of infected persons and social media coverage.

The article's objective is to sum up existing information about social media in the big data environment, and to highlight methods for research and analysis of social media data.

Research results. The principal tool for handling social media data when exploring user behavior in Internet (analysis of responses on messages, shares between friends and groups, analysis of the text part of commentaries, messages, etc.) is appropriately programmed and learned artificial intellect. This refers to the processing of personalized and confidential data that can be structured and analyzed in real time. This type of analysis offers an alternative to traditional studies and an opportunity to have immediate feedback from stakeholders on certain issues, including the ones pertaining to public administration. However, now the overwhelming majority of studies and projections are based on the statistical data available in Internet.

Therefore, two approaches to SMD research and analysis can be highlighted, when SMD defined is either an object of Internet statistics or an object of big data (Figure 1).

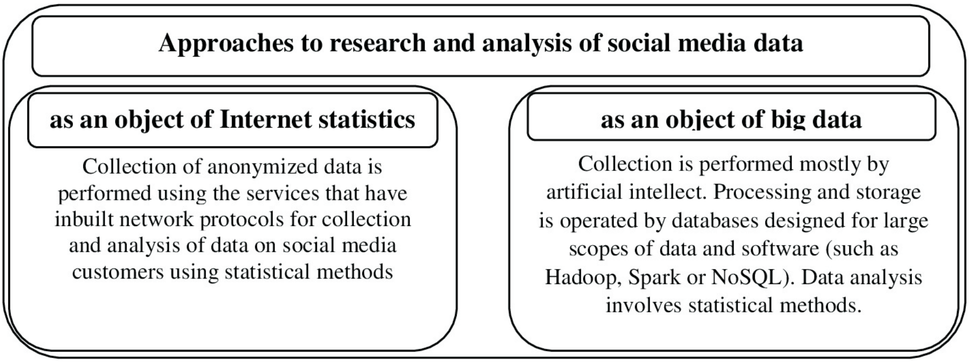
Approaches to research and analysis of social media data
as an object of Internet statistics
Collection of anonymized data is performed using the services that have inbuilt network protocols for collection and analysis of data on social media customers using statistical methods
as an object of big data
Collection is performed mostly by artificial intellect. Processing and storage is operated by databases designed for large scopes of data and software (such as Hadoop, Spark or NoSQL). Data analysis involves statistical methods.

Figure 1. Approaches to research and analysis of social media data

Source: developed by the authors

SMD as an object of Internet statistics are represented by quantitative and qualitative indicators of social media located in Internet.

According to the data of Internet World Stats, Asian residents make the largest segment of Internet users in terms of the population (Figure 2), but it should be noted that in spite of the largest number of Internet users in Asia, social media of Asian origin (such as TikTok) are not on the top in the overall popularity rating.

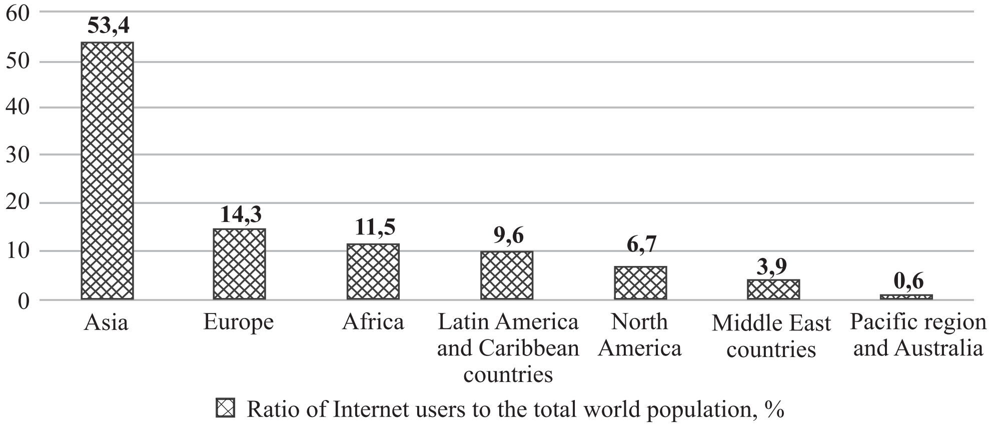

Figure 2. Distribution of Internet users in the world[1]

Source: constructed by the authors using the data from [16]

The data confirming the above said are published on yearly basis by Datareportal (Table 1). It was estimated for 2020 that the total time spend by the humanity in 2020 in the World Web was longer than 1.3 billion years.

Table 1

The most popular social media by number of users, 2020

Serial number	Name of social media	Number of users (billion)	Gender diversity of advertising audience, %	
			Men	Women
1	2	3	4	5
1	Facebook	2.2	56	44
2	YouTube	2.0	54	46
3	WhatsApp	2.0	55	45

[1] Data as of March 31, 2021.

Table 1

1	2	3	4	5
4	Instagram	1.2	49	51
5	Facebook Messenger	1.1	56	44

Source: grouped by the authors on the basis of [17]

Data of Table 1 show that Facebook continues to remain the most popular social media with the total 2.2 billion users, and that the men's registrations in it are 12 percentage points more than women's. By number of users Facebook is followed by YouTube and WhatsApp (2 billion). The fourth most popular social media is Instagram with 1.1 billion of the total users of. It should be noted that Instagram is visually attractive social media, and that is why it is not surprising that its audience mostly consists of women (51%). The list of top ranking social media is closed by Facebook Messenger, an application for message exchange, with the prevailingly male audience (56%) and the total of 1.1 billion users.

It should be noted that Ukraine follows the global trends, as nearly 60%, 43%, and 30% of the Ukrainian Internet users are the users of Facebook, YouTube, and Instagram.

But there are some social media that rapidly gain popularity among users (Table 2). LinkedIn is most often user for setting business contacts, whereas TikTok, a platform for short videos (usually 30 to 15 seconds) has purely entertaining purpose and gained popularity as the pandemic began.

Table 2

Social media gaining popularity, 2020

Serial number	Name of social media	Number of users (million)	Gender diversity of advertising audience, %	
			Men	Men
1	LinkedIn	726.6	57	43
2	TikTok	689.0[2]	51	49
3	Sina Weibo	511.0	53	47
4	Telegram	500.0	59	41
5	Snapchat	498.2	41	59
6	Reddit	430.0	60	40
7	Twitter	353.1	69	31
8	Quora	300.0	59	41
9	Pinterest	200.8	23	77

Source: grouped by the authors on the basis of [17]

Statistical indicators of social media are quite similar. The most common indicators are likes, reposts, comments. The Facebook's toolkit for statistical analysis contains a series of indicators: number of subscribers, number of page views, number of likes, coverage of audience by messages, coverage of stories, interactions with publication (shares, video watches, comments), subscriber characteristics (by age, gender, location). For detailed profile evaluations, indexes of page's love rate, talk rate, amplification rate, conversion or clickability, engagement rate by reach, engagement rate by view etc. are used (Table 3).

Table 3

Statistical indicators for assessment of the audience feedback

Serial number	Indicator	Characterization	Method of estimation
1	2	3	4
1	Love Rate	The message attractiveness for the audience	$\frac{Likes}{Followers} * 100\%$

[2] According to report of "The Verge", TikTok reached 1 billion users in September 2021 [18].

Table 3

1	2	3	4
2	Talk Rate	The message popularity in discussions	$\frac{\textit{Comments}}{\textit{Followers}} * \textit{100\%}$
3	Amplification Rate	The growth in popularity through the audience coverage, thus characterizing the content	$\frac{\textit{Shares}}{\textit{Posts}} * \textit{100\%}$
4	Engagement Volume	An aggregate index of the audience engagement in a page or a message	$\textit{Likes} + \textit{Comments} + \textit{Shares}$
5	Engagement Rate by Reach	The approximate share of those who have seen the publications in a page, and those who have responded on them	$\frac{\textit{NoI}}{\textit{Coverage}} * \textit{100\%}$
6	Daily Engagement Rate	The daily activity of an average statistical subscriber of a page	$\frac{\textit{DER}}{\textit{Number of subscribers}} * \textit{100\%}$
7	Engagement Rate by Message	The message popularity	$\frac{\textit{Coverage of the one post}}{\textit{Number of subscribers}} * \textit{100\%}$
8	Engagement Rate by View	The total number of non-unique views	$\frac{\textit{Number involved} \in 1\,\textit{post}}{\textit{Number of views}} * \textit{100\%}$

Source: constructed by the authors

Because Instagram is most often used for disseminating advertising material, important statistical indicators for this media are number of subscribers and number of views of messages and stories. A specific feature of Twitter is that this social media is a microblogging with user messages not exceeding 280 symbols. In view of this, its key statistical indicators are number of subscribers, number of messages and audience activity. Given that identification of data in twits (themes of discussions or messages) in this media is made mostly by hashtags (using octothorp and keywords), there is a possibility of thematic grouping by emotional coloring of public sentiments.

Social media data as a big data component. Social media analysis has recently become a stable supporting field of research in social and behavioral sciences: this analysis helps define additional influencing factors, trace public sentiments, identify gender diversity with respect to topical themes etc.

Linton C. Freeman argues that only quite recently there appeared established approaches to research and analysis of social media, which formed the general paradigm of SMD research, with analysis (i) motivated by the structural intuition based on users' feedbacks, (ii) built on regular empirical data; (iii) largely supported by graphic images; (iv) based on mathematical and/or computational models [19].

When exploring SMD as a big data component, it should be mentioned that "Microsoft" company has already determined six components characterizing bid data. These components are given by C. Wu, R. Buyya, K. Ramamohanarao:

1. Scopes (scales) of data.
2. Velocity (analysis of flow data).
3. Diversity (various data formats depending on the number of variables).
4. Veracity (or accuracy, with emphasis on the reliability of data sources).
5. Variability (complexity of a data set).
6. Visibility (to take an informative decision, it is necessary to have all the required data) [20].

Therefore, SMD can be given a meaningful characterization using these components (Table 4).

Table 4

The conformity of social media data with big data components

Component	Characterization
Scopes	SMD demonstrate annual exponential growth. Nowadays[3], the global number of social media users is 19 billion more than in the last quarter of 2020, with 2.5^0 quintillion bytes of data created each day[4]
Velocity	Each second users send 575,000 twits to Twitter, display 67,000 photographs in Instagram; 510,000 comments are published and 542,000 statuses are updated in Facebook. Each day 240,000 photographs are downloaded in Facebook. Users of TikTok view 167,000,000 videos each second[5]
Diversity	Each social media has its own content specifics (various image and video formats, video duration, requirements on confidentiality etc.)
Veracity	There exist several pressures on the veracity: - social media platforms have created the reality distortion field, i. e. the field where users quite often give out desirable for valid (especially in Instagram); - designers of social media platforms are not interested in providing and processing of accurate data; - language inconformity; lack of reliable software for SMD processing
Variability	The inaccuracy sometimes may be astonishing, because an immense part of the information in social media is collected through crowdsourcing
Visibility	To increase the trust for SMD, platform designers need to improve their computing infrastructure and, probably, reduce their functionality. However, given proper collection, processing and analysis, SMD can be useful for making original judgements in some thematic fields

Source: grouped by the authors by data from [21–23]

Results of Table 4 allow to highlight key problems faced at early phases of work with SMD: lack of SMD management policy; gradually increasing scopes of SMD; lack of reliable software and tools for collection, processing, analysis and storage of SMD; inconsistency of SMD; inaccuracy, unreliability and futility of some SMD.

Based on the argument that “many big data projects are technology-driven and thus, expensive and inefficient. It is often unclear how to exploit existing data resources and map data, systems and analytics results to actual use cases”, M. A. Kaufmann proposed his own practical method for big data processing: BDM$^{\text{cube}}$ model. This method is quite logical and acceptable for SMD processing as well. It involves data preparation, analysis, interaction, effectuation, intelligence, and the effective management of new knowledge in this process, M. A. Kaufmann argued that “…Data intelligence is a knowledge-driven cross-platform function that ensures that these knowledge assets are optimally deployed, distributed, and utilized over all layers of BDM” [24].

Conclusions. As shown by the study, SMD are a type of data which specificity is attributable to their inconsistency, lack of structuring and rapid growth. Theoretical and applied research of SMD focused on their collection, processing and sound analysis has not been sufficient by now, which create barriers in understanding the importance of using these data. The study revealed several problems which solutions would require, apart from a multifaceted and complex approach to collection and processing, highly competent teams of specialists in various subject fields, including experts in computations, experts in machine learning and statisticians.

The problems occurring now in SMD handling can be explained by the novelty of data. Yet, they make one realize that SMD involve an absolutely new dimension of scopes, velocity, diversity, veracity, variability and visibility, and just like big data, they require unconventional and innovative methods for their analysis and processing.

[3] Estimates available in July 2021.

[4] Estimates available in the end of 2017.

[5] Estimates available in the end of 2020.

References

1. Onlain 2020. Yak pandemiia vplynula na onlain-korystuvannia [Online 2020. What was the pandemic impact on online use]. Suspilne, October 29, 2020. Retrieved from https://suspilne.media/74631-onlajn-2020-ak-pandemia-vplinula-na-onlajn-koristuvanna/ [in |Ukrainian].
2. Schwaiger J., Hammerl T., Florian J., & Leist S. (2021). UR: SMART–A tool for analyzing social media content. Information Systems and e-Business Management, 19, 1275–1320. https://doi.org/10.1007/s10257-021-00541-4
3. Sarprasatham M. (2016). Big Data in Social Media Environment: A Business Perspective. Social Media Listening and Monitoring for Business Applications, pp. 70–93. https://doi.org/10.4018/978-1-5225-0846-5.ch004
4. Bing L., Chan K. C. C., & Ou C. (2014). Public sentiment analysis in Twitter data for prediction of a company's stock price movements. 11th IEEE International Conference on e-Business Engineering (ICEBE 2014). Sun Yat-sen University, Guangzhou, China. Retrieved from https://research.tilburguniversity.edu/en/publications/public-sentiment-analysis-in-twitter-data-for-prediction-of-a-com
5. Budiharto W., & Meiliana M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. Journal of Big Data, 5, 51. https://doi.org/10.1186/s40537-018-0164-1
6. Seltzer E. K., Horst-Martz E., Lu M., & Merchant R. M. (September, 2017). Public sentiment and discourse about Zika virus on Instagram. Public Health, 150, 170–175. https://doi.org/10.1016/j.puhe.2017.07.015
7. Ahmad T., Alvi A., & Ittefaq M. (2019). The use of social media on political participation among university students: an analysis of survey results from rural Pakistan. SAGE Open, July-September, 1–9. https://doi.org/10.1177/2158244019864484
8. Svidronova M., Kascakova A., & Bambusekova G. (2019). Social media in the presidential election campaign: Slovakia 2019. Administratie si Management Public, 33, 181–194.
9. Mayfield A. (2008). What is social media? Vol. 1.4. Retrieved from: https://tavaana.org/sites/default/files/what-is-social-media-uk.pdf
10. Zernetska O. (2009). Hlobalna politychna blohosfera – nova arena politychnoi komunikatsii. Politychnyi menedzhment [The global political blogosphere: a new arena of political communication]. Politychnyi menedzhment – Political Management, 2, 13–26. Retrieved from http://dspace.nbuv.gov.ua/bitstream/handle/123456789/59793/02-Zernetska.pdf?sequence=1 [in Ukrainian].
11. McAllister I. (2016). Internet use, political knowledge and youth electoral participation in Australia. Journal of Youth Studies, vol. 19, issue 9, 1220–1236. https://doi.org/10.1080/13676261.2016.1154936
12. World Health Organization. (September, 2020). Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC. Retrieved from https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation
13. World Health Organization. (2020). WHA 73. Retrieved from https://apps.who.int/gb/e/e_wha73.html
14. Alton M. K. Chew, & Gunasekeran D. V. (2021). Social Media Big Data: The Good, The Bad, and the Ugly (Un)truths. Front. Big Data, 4, 623794. https://doi.org/10.3389/fdata.2021.623794
15. Domalewska D. (2021). An analysis of COVID 19 economic measures and attitudes: evidence from social media mining. Journal of Big Data, 8, 42. https://doi.org/10.1186/s40537-021-00431-z
16. Internet World Stat. Retrieved from https://www.internetworldstats.com/
17. Digital 2021: Global Owerview Report (January, 2021). Retrieved from https://datareportal.com/reports/digital-2021-global-overview-report
18. TikTok says it has passed 1 billion users. (2021). The Verge. Retrieved from https://www.theverge.com/2021/9/27/22696281/tiktok-1-billion-users

19. Freeman L. C. (2004). The Development of social network analysis a study in the sociology of science. Empirical Press Vancouver, BC Canada, pp. 3–4. Retrieved from https://www.researchgate.net/profile/Linton-Freeman-2/publication/239228599_The_Development_of_Social_Network_Analysis/links/54415c650cf2e6f0c0f616a8/The-Development-of-Social-Network-Analysis.pdf
20. Wu C., Buyya R., & Ramamohanarao K. (2016). Chapter 1. Big Data Analytics = Machine Learning + Cloud Computing. Big Data Principles and Paradigms. R. Buyya, R. N. Calheiros, A. V. Dastjerdi (Eds.). Elsevier, pp. 3–38. https://doi.org/10.1016/B978-0-12-805394-2.00001-5
21. Kharkovchuk O. Dynamika zrostannia audytorii sotsialnykh merezh: porivniuiemo kvartalni zvity DataReportal za 2020 i 2021 roky [Dynamics of growth in the audience of social networks: comparing quarterly reports for 2020 and 2021]. Webpromo, July 8, 2021. Retrieved from https://web-promo.ua/ua/blog/dinamika-rosta-auditorii-socialnyh-setej-cravnivaem-kvartalnye-otchety-datareportal-za-2020-i-2021-gody/ [in Ukrainian].
22. Marr B. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Bernard Marr & Co. Retrieved from https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/
23. Data Never Sleeps: Here’s What Happens Every Minute Online [Infographic]. (October, 2021). Retrieved from https://hcsmmonitor.com/2021/10/05/data-never-sleeps-heres-what-happens-every-minute-online-infographic/
24. Kaufmann M. A. (2019). Big Data Management Canvas: A Reference Model for Value Creation from Data. Big Data Cogn. Comput., 3(1), 19. https://doi.org/10.3390/bdcc3010019

Bibliographic description for quoting:

Osaulenko O., Horobets O. (2021). Social Media Data in the Big Data Environment. *Scientific Bulletin of National Academy of Statistics, Accounting and Audit – Naukovyy visnyk Natsionalnoi akademiyi statystyky, obliku ta audytu*, 3-4, 23-31. doi:10.31767/nasoa.3-4-2021.03.