DOI: 10.34616/150561

Dr Jarosław Greser
Politechnika Warszawska, Wydział Nauk Społecznych i Administracji
Vrije Universiteit Brussel, Cyber and Data Security Lab
ORCID: 0000-0002-1021-6142

Access to health data for scientific research. Remarks in the light of Article 40 DSA

Abstract

The primary objective of the Digital Services Act is to enhance the transparency of the operations of VLOPs and VLOSEs. This is achieved by ensuring the possibility of effective control by public entities and by preventing negative phenomena covered by systemic risks through research by independent researchers. However, the data that VLOPs and VLOSEs collect as part of their services can also be used for research for public good. This is particularly relevant in the case of information on the health status of their users. This data can be important in the development of new health products and services and for the planning of public policies in this area. This article explores whether the access to data provided for in Article 40 of the Digital Services Act allows for research in health-related areas.

The article is divided into four parts. The first part examines the concept of health data, and the second considers whether medical data falls within the purposes of allowing data sharing. The third part of the article analyses the conditions that limit the release of data for research purposes. The final part of the article contains conclusions andpoints out that the Digital Services Act allows health information to be made available for scientific research in this area, but that this is very limited by the conditions contained in the act. However, the interpretation of these conditions can vary, and it largely determines how much and what data researchers can receive in practice. Consequently, the role of the Digital Services Coordinators is of great importance in shaping the practice of fulfilling data access requests.

Keywords

health data, open science, VLOP, VLOSE, Digital Service Act, access to data

The role of user-generated data in medical research

The slogan ‘data is the new oil’ aptly reflects its importance in modern society. The development of the internet has led to an unprecedented opportunity in human history to generate and analyse data. According to some forecasts, the total amount of data generated by humans will reach 175 zettabytes per year by 2025, an increase of 530% compared to 2018[1], which seems to be a significant underestimate. The main sources

[1] *Europejska strategia w zakresie danych. Prognozy na 2025 r.* https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_pl [access 1.03.2024].

are the Internet of Things[2] and online platforms. In the case of the latter, the scale can be illustrated by the example of Facebook, which generates 510,000 comments and 136,000 photos every minute[3], totalling 4 petabytes of data per day[4].

This data includes various pieces of information that can be used in medical research, such as eating habits, well-being, side effects of drugs or medical devices, and treatment effectiveness[5]. This information offers more alternatives compared to traditional research that relied solely on medical records[6]. In the era of machine learning, opportunities have arisen to analyse data more accurately and combine it with information from other sources. This can lead to more precise analyses[7] and a better understanding of the underlying structure and patterns in the data[8]. This has practical applications in medical research, including the treatment of chronic diseases[9], improving the functioning of patients with neurodegenerative diseases[10], predicting the occurrence and spread of infectious diseases[11], and finding new ways to treat mental illness[12]. Additionally, it may have wide applications in evidence-based policy-making in such areas as national defence, public health, tackling climate change and improving public services[13].

[2] J. Greser, *Wybrane problemy funkcjonowania Internetu Rzeczy w relacji do praw człowieka*, [in:] B. Gronowska, P. Sadowski (red.), *25-lecie wejścia w życie Europejskiej Konwencji Praw Człowieka w Polsce*, Toruń 2022.

[3] O. Maddy, *Wild and Interesting Facebook Statistics and Facts*, https://kinsta.com/blog/facebook-statistics/ [access 1.03.2024].

[4] *Facebook Research*, https://research.facebook.com/blog/facebook-s-top-open-data-problems/ [dostęp 1.03.2024].

[5] J. Greser, *Etyczne problemy wdrażania medycznego Internetu Rzeczy*, „Prawo Mediów Elektronicznych" 2020 nr 3, p. 4–5.

[6] R.I. Horwitz, *Comparison of epidemiologic data from multiple sources*, „Journal of Chronic Diseases" 1986, no 11, p. 889-896.

[7] N. Peek, J.H. Holmes, J. Sun, *Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics*, „Yearbook Medical Informatics" 2014, no 1, p. 42-47.

[8] L. Brady, W. Ting, *Chatting about ChatGPT: how may AI and GPT impact academia and libraries?,* „Library Hi Tech News Volume" 2023, no 3, p. 26.

[9] D. Su *et al.*, *Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials*, „Diabetes Research and Clinical Practice" 2016, no 116, p. 136-148.

[10] J. Greser, *A step forward in health-related IoT cybersecurity: remarks on the proposal for a liability for defective products directive*, „Frontiers Digital Health" 2023, no 5, https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1193255/full [access 1.03.2024].

[11] S. Yanga, M. Santillanab, S. C. Koua, A*ccurate estimation of influenza epidemics using Google search data via ARGO*, „Proceedings of the National Academy of Sciences" 2015, no 47, p. 14473-14478.

[12] J. Ive *Leveraging the potential of synthetic text for AI in mental healthcare*, „Frontiers Digital Health" 2022, no 4, https://doi.org/10.3389/fdgth.2022.1010202 [access 1.03.2024].

[13] E. Hazelkorn, A. Gibson, *Public goods and public policy: what is public good, and who and what decides*?, „Higher Education" 2019, no 78, p. 260.

Access to data collected by private entities is often restricted due to the assumption that the data becomes the private property of the company upon collection[14]. As a result, the data falls under the protection of private law provisions such as trade secrets or database protection, and is excluded from the framework of provisions allowing access and re-use of public data[15]. This makes it very challenging, and in many cases impossible, to conduct research based on such data. However, new legal frameworks are emerging at the European Union level that may allow greater access to data held by private parties. One such framework is the European Health Data Space[16], which is yet to be implemented. In the case of IoT devices, the Data Act[17] regulates access to non-personal data and will play a significant role. The Digital Service Act[18], one of the initiatives regulating the digital market in the EU, plays a crucial role for online platforms.

The purpose of this article is to examine the extent to which health data can be shared in compliance with Article 40 of the DSA, which establishes the normative framework for data sharing by online platforms. The article is divided into four sections. The first part analyses the concept of health data, followed by an analysis of whether medical data falls within the scope of purposes that allow data sharing. Thirdly, other conditions for data sharing outlined in the Digital Service Act are explored. The final part presents conclusions and recommendations.

The concept of health data

To define the concept of health data, it is necessary to refer to the General Data Protection Regulation (GDPR)[19]. This regulation provides a legal definition of 'per-

[14] P. Kyung, *Data as Public Goods or Private Properties?: A Way Out of Conflict Between Data Protection and Free Speech*, „Irvine Journal of International, Transnational, and Comparative Law" 2021, no 77, https://scholarship.law.uci.edu/ucijil/vol6/iss1/5 [access 1.03.2024].

[15] Communication from The Commission A European strategy for data, COM(2020) 66 final, p. 7.

[16] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM/2022/197 final.

[17] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), Dz. Urz. UE L 2023/2854 from 22.12.2023.

[18] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), Dz. Urz. UE L 277 from 27.10.2022.

[19] Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Dz. Urz. UE L 119 from 4.5.2016.

sonal data', 'data concerning health', and related 'biometric data' and 'genetic data', which are subsets of 'personal data'.

According to Article 4(1) of the GDPR, personal data means any information relating to an identified or identifiable natural person. It is crucial to note that any information, regardless of its expression, can be considered personal data[20]. However, there is a tendency to classify information too broadly as personal data[21]. In practice, it can be challenging to distinguish between personal and non-personal information, especially when dealing with data collected from online platforms. When in doubt, it is advisable to apply the data protection regulations to such data, especially when personal and non-personal data sets are linked[22]. This provision enhances the legal certainty for researchers while also enabling access to data.

It is important to note that article 40(8)(d) of the DSA clearly states that both personal and non-personal data may be accessed, subject to specific data security and confidentiality requirements for each type of data. Additional conditions must be met for the former, which are directly derived from the data protection legislation. This includes Article 40(8)(g), which concerns the protection of data cited in the publication of research results. Therefore, researchers do not have to limit their research to just one type of data, nor is applying for personal data the sole reason for not gaining access.

The term 'data concerning health' is equally broad in scope. Pursuant to Article 4 (15) of GDPR, it refers to personal data that pertains to the physical or mental health of an individual, including the provision of healthcare services, and which discloses information about their health status. Recital 35 GDPR provides examples of such data, including information related to the registration and provision of healthcare services, as well as any information regarding a disease, disability, disease risk, medical history, clinical treatment, or the physiological or biomedical state of the data subject. Furthermore, the recital highlights that the information pertains to the data subject's physical or mental health status in the past, present, or future, and can originate from various sources, including online platforms. The European Court of Justice[23] sup-

[20] P. Litwinski, *Komentarz do art. 4*, [in:] P. Litwinski (red.), *Ogólne rozporządzenie o ochronie danych osobowych. Ustawa o ochronie danych osobowych. Wybrane przepisy sektorowe. Komentarz*, Warszawa 2018, Legalis, NB 4.

[21] N. Purtova, *The law of everything. Broad concept of personal data and future of EU data protection law*, „ Law, Innovation and Technology" 2018, no 1.

[22] L. Brygave, L. Tossoni, *Commentary to article 4*, [in:] Ch. Kuner *et al.* (red.), *The EU General Data Protection Regulation (GDPR). A commentary*, Oxford 2020, p. 113

[23] European Court of Justice judgment of 6.11.2033 case C-101/01, *Criminal proceedings against Bodil Lindqvist*, European Court Reports 2003, I-12971.

ports the trend towards a comprehensive interpretation of health data, as evidenced by its recognition of sick leave as such information. It is important to note that, under Article 9(1) of the GDPR, data concerning health is considered a special category of personal data. From the point of view of the requirements under the DSA, this will be relevant in determining the higher level of protection of the data received, which must be provided by the researcher, and the burden of proof that he or she is able to do so lies with the researcher[24].

Health data may also encompass 'biometric data' and 'genetic data'. Biometric data refers to the physical, physiological, or behavioural characteristics of a natural person which allow or confirm the unique identification of that natural person. These include, but are not limited to, the characteristics of the iris of the eye, the colour of the voice, the pattern of blood vessels. Furthermore, the literature indicates that biometric data encompasses behavioural characteristics, including the manner in which a person moves[25]. Generally, biometric data remains constant throughout an individual's life[26]. Biometric data is obtained through technical methods that allow for the unique identification of an individual. This may include data related to health status, but further analysis is required in this area. The literature suggests that processing biometric data carries the risk of discriminatory misidentifications[27], bias[28], or other violations of individual rights[29]. It is important to note that the DSA does not impose an obligation to process data in a way that avoids these issues. Therefore, the mere suspicion of such conduct is not sufficient grounds for denying access to data. However, such actions are in clear violation of research ethics, and may breach national and EU laws and results obtained through such means cannot be considered reliable.

When it comes to genetic data, the matter is more complicated. According to Article 4(13) GDPR, this includes 'personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from

[24] J. Greser, *Zasady dostępu i wykorzystania danych posiadanych przez VLOP i VLOSE przez Komisję Europejską i Koordynatorów do spraw usług cyfrowych*, „Prawo Nowych Technologii" 2023, no 3-4, p. 147.

[25] P. Litwinski, *Komentarz do art. 4…*, NB 137.

[26] L. Brygave, L. Tossoni, *op. cit.*, p. 208.

[27] J. Buolamwini, T. Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, „Proceedings of Machine Learning Research" 2018, no 81, p. 1-15.

[28] C. Schwemmer, C. Knight, E. D. Bello-Pardo, S. Oklobdzija, M. Schoonvelde, J. W. Lockhart, *Diagnosing Gender Bias in Image Recognition Systems*, „Socius" 2020, no 6, p. 1-17.

[29] V. Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*, New York 2018, p. 17.

an analysis of a biological sample from the natural person in question'. One of the conditions for classifying information as genetic data is that it provides unique information about the individual's health[30], which automatically categorises it as health data[31]. Although there is no automatic classification for physiological traits, they may also fall under the scope of health data due to the broad definition. It is important to note that obtaining such information requires specialised technical means. However, recital 34 GDPR states that any means leading to equivalent information from DNA or RNA samples can be used. As an example, it is possible to obtain genetic data through image analysis[32]. It is possible that the platform may process genetic test of images posted by users for instance on social media. This is especially true since genetic tests have applications beyond medical purposes. It is important to note that if the analysis does not pertain to physiology or health, the data will not be considered genetic data[33].

General rules on access to data for scientific purposes

Article 40 of the Digital Services Act outlines the principles for data sharing. These principles can be divided into three groups: defining who can receive and share data, specifying what data can be transferred, and establishing the procedure for requesting and terminating access to data. This article will focus on the first two premises, which are central to the problem under analysis.

Scope of subject matter

Under Article 40, the only entities obliged to share data are very large online platforms (VLOP) and very large online search engines (VLOSE). The criteria for determining these categories are defined in Article 33 of the DSA and are based on the ratio of users to the population in the EU. The Commission designated 23 VLOPs and VLOSEs. The first 19 entities were designated on 25 April 2023[34],

[30] P. Litwinski, *Komentarz do art. 4, op. cit*., NB 132.

[31] L. Brygave, L. Tossoni, *op. cit*., p. 203.

[32] Y. Gurovich *et al.*, *Identifying facial phenotypes of genetic disorders using deep learning,* „Nature Medicine" 2019, no 25, p. 60–64.

[33] L. Brygave, L. Tossoni, *op. cit.,* p. 202.

[34] *Digital Services Act: Commission designates first set of Very Large Online Platforms and Search Engines*, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413 [access 1.03.2024].

three additional VLOPs were designated on 20 December 2023[35], and the last one was designated on 26 April 2024[36]. Two of designated VLOP have appealed against the Commission's decision to the ECJ[37]. Given the criteria for recognition as VLOP and VLOSE, no significant change in the number of compliant companies is expected. Such a regime excludes other online platforms, including those that are independent entities under commercial law but under the effective control of VLOP and VLOSE. This could include start-ups or companies that benefit from preferential access to data, computing power or infrastructure in exchange for services to these entities. This appears to be a loophole that creates the possibility of circumventing the provisions of the Regulation, despite the fact that the issue of attributing liability to the entity exercising effective control has been advocated in both public[38] and private law[39] doctrine for many years. This should be seen as a significant shortcoming of the Regulation. It seems that a better criterion for assessing whether a platform should exchange information is the systemic risk it poses. The analysis of DSA recitals 76 and 79 shows that this concept has been used to justify the automatic inclusion of VLOP and VLOSE in additional obligations. At the same time, there is no tool to examine whether such risks are created by other actors.

Those who may request access to data are divided into two groups: vetted researchers referred to in Article 40(8) and researchers referred to in Article 40(12). For the first group, the decision to grant this status is taken by the Digital Services Coordinator of the institution, after a number of conditions set out in this article have been met. These include being part of a research organisation, acting independently from commercial interests, disclosing the funders of the research, being able to meet specific data security and confidentiality requirements, demonstrating that the research is carried out for the purposes set out in Article 40(4) and that the results will contribute to the achievement of those purposes, using the necessary and proportionate amount of data to carry it out, and making the results of its research publicly available free

[35] Supervision of the designated very large online platforms and search engines under DSA, https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses [access 1.03.2024].

[36] *Commission designates Shein as Very Large Online Platform under the Digital Services Act*, https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2326 [access 30.04.2024].

[37] J. Tar, *Amazon joins Zalando in challenging very large online platform designation*, https://www.euractiv.com/section/platforms/news/amazon-joins-zalando-in-challenging-very-large-online-platform-designation/ [access 1.03.2024].

[38] A. Clapham, S. Jerbi, *Categories of Corporate Complicity in Human Rights Abuses*, „Hastings International and Comparative Law Review" 2001, no 24, p. 339-349.

[39] P. Muchlinski, *Multinational Enterprises and the Law*, Blackwell 1999, p. 336-338.

of charge. The cumulative fulfilment of these conditions is required, which may be difficult given their evaluative nature and uncertainties in their interpretation[40]. At the same time, achieving this status is rewarded by the opportunity to access data that is not publicly available and is legally protected, for example, by trade secret or personal data protection rules. In addition, they may carry out research for the detection, identification and understanding of systemic risks in the Union, as referred to in Article 34(1) DSA, and for the assessment of the adequacy, effectiveness and impact of risk mitigation measures, as referred to in Article 35 DSA.

The decision to grant the status of researcher is a matter for the VLOP or VLOSE, which will independently assess the fulfilment of the conditions set out in Article 40(12). These include independence from commercial interests, disclosure of the funders of the research, ability to meet the specific data security and confidentiality requirements, demonstration that the research is carried out for the sole purpose of detecting, identifying and understanding systemic risk in the Union, as referred to in Article 34(1), and that the results obtained will contribute to the achievement of that objective, using the necessary and proportionate amount of data to carry out the research. Fewer requirements to be met will result in limited access to information, which only cover data publicly available in the VLOP or VLOSE online interface. However, the data should be available in real time. It should be noted that both the wording of the provision itself and recital 97 DSA indicate that the status of researcher can be obtained by individuals who are not affiliated to a research organisation. This leaves considerable scope for the development of research within citizen science[41].

Scope of data shared

The authors of the DSA have opted for an open catalogue of data that can be requested by researchers, but the scope of which is determined by the purposes for which they are to be used. As mentioned above, they relate to the issues identified in Articles 34 and 35 of the DSA. From the perspective of access to health information, the key question is whether it falls within the scope of systemic risk, which is a central concept used in these provisions. It should be noted that it is not defined, so it is necessary to refer to the definitions developed by researchers specialising in risk studies. According

[40] J. Greser, *Access to data for academic purposes under the Digital Services Act*, „Przegląd Europejski" [in print].

[41] B. Balázs *et. al.*, *Data Quality in Citizen Science*, [w:] K. Vohland *et al.* (red.), *The science of citizen science,* Berlin 2021, p. 139-199.

to the literature, systemic risk refers to “risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by co-movements (correlation) among most or all parts”[42]. It also has the potential for a threat or hazard to propagate disruptions or losses to multiple interconnected parts of complex systems[43]. It is used in a similar context in legislation regulating the financial system, for example in Article 2(c) of Regulation 1092/2010[44]. In view of this and the content of the recitals of the DSA, it must be assumed that the analysis goes far beyond effects that are merely infringements of the law. It is also necessary to analyse the legal content that may increase the level of systemic risk[45].

In addition, Article 34(1) identifies the areas of systemic risk to be taken into account in the risk analysis. In addition to Article 34(1)(d), issues related to the health of individuals have not been directly identified, but the interpretation of the provisions makes it possible to identify areas that will touch on it. The first is the risks associated with the dissemination of illegal content through VLOP and VLOSE services. In particular, this will concern matters that are explicitly prohibited by criminal law of individual Member States. These include offers for the sale of counterfeit medicines or information on the production of medicines. At the same time, it appears that misinformation and fake news will not be covered by this provision, as long as they are not prohibited by national law.

The second area is the risk of ‘actual or foreseeable negative effects for the exercise of fundamental rights’. In the Charter of Fundamental Rights[46] we find references to health in Article 3, which refers to the right to the integrity of the person, in Article 31 on fair and just working conditions, and in Article 35, which states that a high level of human health protection shall be ensured. It should be noted that this area is very broad and will include other rights contained in the Charter, in line with the concept of interdependence of human rights[47].

[42] G. Kaufman, S. Kenneth, *What is systemic risk, and do bank regulators retard or contribute to it?,* „Independent Review” 2003, no 7, p. 372.

[43] O. Renn *et al.*, *Systemic Risks from Different Perspectives*, „Risk Analysis” 2022, no 42, p. 1903, https://doi.org/10.1111/risa.13657 [access 1.03.2024]; International Risk Governance Council, *Guidelines for the governance of systemic risks*, Lausanne 2018.

[44] Regulation (EU) No 1092/2010 of the European Parliament and of the Council of 24 November 2010 on European Union macro-prudential oversight of the financial system and establishing a European Systemic Risk Board, Dz. Urz. UE L 331 from 15.12.2010.

[45] P. Podrecki, *Komentarz do art. 34*, [in:] M. Grochowski (red.), *Rynek cyfrowy. Akt o usługach cyfrowych. Akt o rynkach cyfrowych. Rozporządzenie platform-to-business. Komentarz*, Warszawa 2024, NB 10.

[46] Charter of Fundamental Rights of the European Union, Dz. Urz. UE C 326/02 from 26.10.2012.

[47] M. Nowak, *Introduction to the International Human Rights Regime*, The Hague 2003, p. 70.

As mentioned above, Article 34(1)(d) requires the assessment of risks associated with serious negative consequences for the physical and mental well-being of the person. As stated in the literature, the potential harm to mental health is significant[48]. It is important to note the evaluative nature of the term "serious". It seems to be interpreted as referring to effects that are occurring now. However, it is worth highlighting that many services provided by online platforms are reported to have a negative impact on mental health in the distant future[49].

A separate area is that of public health information. This concept has no legal definition and is said to cover a wide range of activities such as education, promotion of healthy lifestyles, research into the prevention of disease and injury, and the detection, prevention and response to infectious diseases[50]. It should be noted that, according to Article 4(2)(k) of the Treaty on the Functioning of the European Union, public health is a shared competence between the Union and the Member States. As explained in the doctrine, this makes it possible to distinguish the following areas of Union action: patients' rights in cross-border healthcare, including the coordination of social security systems; combating health risks, including those related to communicable diseases; blood transfusions; transplantation of human tissues and organs; control of the quality and safety of medicinal products, medical devices and cosmetics; mutual recognition of professional qualifications, including in the health professions; biotechnologies and cell banks; promotion of healthy lifestyles[51]. It is therefore reasonable to assume that at least these issues should be covered by the systemic risk analysis carried out by VLOP and VLOSE, and thus fall within the scope of data to which vetted researchers can apply for access. It also seems that such a broad public health perspective provides a basis for analyses of misinformation and fake news, all the more so as they have very acute effects in this area and online services, especially social media, are one of the main sources of their dissemination[52].

[48] J. Gao *et al.*, *Mental health problems and social media exposure during COVID-19 outbreak*, „PLOS one", 2020, https://doi.org/10.1371/journal.pone.0231924 [access 1.03.2024].

[49] S. Chancellor, M. De Choudhury, *Methods in predictive techniques for mental health status on social media: a critical review,* „npj Digital Medcine" 2020, no 3, https://doi.org/10.1038/s41746-020-0233-7 [access 1.03.2024].

[50] *What is public health?,* https://www.hsph.harvard.edu/communications-guide/what-is-public-health/ [access 1.03.2024].

[51] J. Barcik, *Odpowiedzialność publicznoprawna*, [in:] A. Barczak-Oplustil, T. Sroka (red.), *System Prawa Medycznego,* vol. 6, Warszawa 2023, p. 194-195.

[52] C. Melchior, M. Oliveira, *Health-related fake news on social media platforms: A systematic literature review*, „New Media&Society" 2021, no 6, https://journals.sagepub.com/doi/abs/10.1177/14614448211038762 [access 1.03.2024].

In the case of Article 35 DSA on risk reduction measures, it is not possible to distinguish directly between areas relating to individual health and those relating to public health. However, this does not mean that such information is not available. On the contrary, it seems that any of the measures mentioned in this article may apply to a health-related area. In the case of this provision, the focus is on the functional aspect and the effectiveness of the solutions used by VLOP and VLOSE to reduce the risks defined in Article 34 of DSA. Therefore, the scope of the shared data will include all the information that can be requested under this Article.

Conclusions

There is no doubt that the data collected by online platforms is a very valuable source of analysis for researchers working in both individual and public health. It can provide new insights or complement conclusions drawn from other information. Its importance is growing with the development of machine learning algorithms and their increasing use in medicine. At the same time, access to information held by private parties is very limited. Against this background, the Digital Services Act was passed, which explicitly formulates rules for data sharing. The analysis of the DSA allows us to answer the first part of the research question posed in the introduction in the affirmative, as the DSA allows health information collected by online platforms to be extracted and used for research purposes. This should be seen as a step in the right direction and a clearly positive solution that will benefit society as a whole.

On the other hand, when assessing the scope of the data to be exchanged by VLOP and VLOSE, it should be noted that their scope will cover all data, i.e. both personal and non-personal data, and in the case of the former, also biometric and genetic data. The absence of restrictions in this respect is to be well appreciated. Similarly, the requirement for researchers to meet certain criteria for the protection of such data should be welcomed. However, care must be taken to ensure that this requirement does not become a barrier to access. In this respect, the practice of the Digital Services Coordinator of Establishment for vetted researchers and VLOP and VLOSE for researchers will play an important role.

On the other hand, the restriction of data sharing to VLOP and VLOSE can only be assessed negatively. Especially in the case of health research, it could be useful for researchers to obtain information from online platforms that do not have this status. These could be, for example, blog sites or specialised online forums. Similarly, the vague requirements for obtaining the status of vetted researcher and researcher and, in the case of the latter, the restriction on the scope of the research conducted, should be evaluated. Nevertheless, compared to the legal situation prior to the adoption of the DSA, at least we have a defined circle of institutions that can submit applications and a procedure for their examination by bodies independent of VLOP and VLOSE, which is a significant step forward.

Regarding the scope of data to be shared, the design of the legislation allows for both individual and public health information. Limiting the scope by specifying objectives related to research and systemic risk mitigation does not seem particularly restrictive if a research-friendly interpretation is adopted. This is particularly the case given the broad scope of the rights covered by the Charter of Fundamental Rights and the variety of aspects covered by public health.

In conclusion, the Digital Service Act is an important step towards making privately collected data more accessible. Its adoption and consideration should be seen as clearly positive. However, it is important to highlight the limitations that arise from it. These are the result of the area it regulates, i.e. the creation of a safe, predictable and trustworthy online environment, and not the direct regulation of access to data. However, it seems to open the way for further regulations that will allow access to data by more actors and for more purposes.

Bibliography

Balázs B. *et. al.*, *Data Quality in Citizen Science*, [in:] K. Vohland *et al.* (red.), *The science of citizen science,* Berlin 2021.

Barcik J., *Odpowiedzialność publicznoprawna*, [in:] A. Barczak-Oplustil, T. Sroka (red.), *System Prawa Medycznego,* vol. 6, Warszawa 2023.

Brady L., Ting W., *Chatting about ChatGPT: how may AI and GPT impact academia and libraries?,* „Library Hi Tech News Volume" 2023, no 3, p. 26.

Brygave L., Tossoni L., *Commentary to article 4*, [in:] Ch. Kuner *et al.* (red.), *The EU General Data Protection Regulation (GDPR). A commentary*, Oxford 2020.

Buolamwini J., Gebru T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, „Proceedings of Machine Learning Research" 2018, no 81.

Chancellor S., De Choudhury M., *Methods in predictive techniques for mental health status on social media: a critical review,* „npj Digital Medcine" 2020, no 3.

Clapham A., Jerbi S., *Categories of Corporate Complicity in Human Rights Abuses*, „Hastings International and Comparative Law Review" 2001, no 24.

Eubanks V., *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*, New York 2018.

Facebook Research, https://research.facebook.com/blog/facebook-s-top-open-data-problems/.

Gao J. *et al.*, *Mental health problems and social media exposure during COVID-19 outbreak*, „PLOS one" 2020, https://doi.org/10.1371/journal.pone.0231924.

Greser J., *A step forward in health-related IoT cybersecurity: remarks on the proposal for a liability for defective products directive*, „Frontiers Digital Health" 2023, no 5https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1193255/full.

Greser J., *Access to data for academic purposes under the Digital Services Act*, „Przegląd Europejski" (in printing).

Greser J., *Etyczne problemy wdrażania medycznego Internetu Rzeczy,* „Prawo Mediów Elektronicznych" 2020 nr 3, p. 4–5.

Greser J., *Wybrane problemy funkcjonowania Internetu Rzeczy w relacji do praw człowieka*, [in:] B. Gronowska, P. Sadowski (red.), *25-lecie wejścia w życie Europejskiej Konwencji Praw Człowieka w Polsce*, Toruń 2022.

Greser J., *Zasady dostępu i wykorzystania danych posiadanych przez VLOP i VLOSE przez Komisję Europejską i Koordynatorów do spraw usług cyfrowych*, „Prawo Nowych Technologii" 2023, no 3-4.

Gurovich Y. *et al.*, *Identifying facial phenotypes of genetic disorders using deep learning,* „Nature Medicine" 2019, no 25.

Hazelkorn E., Gibson A., *Public goods and public policy: what is public good, and who and what decides*?, „Higher Education" 2019, no 78.

Horwitz R.I., *Comparison of epidemiologic data from multiple sources*, „Journal of Chronic Diseases" 1986, no 11, p. 889-896.

Europejska strategia w zakresie danych. Prognozy na 2025 r., https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_pl.

International Risk Governance Council, *Guidelines for the governance of systemic risks*, Lausanne 2018 International Risk Governance Council , *Guidelines for the governance of systemic risks*, Lausanne 2018.

Ive J., *Leveraging the potential of synthetic text for AI in mental healthcare*, „Frontiers Digital Health" 2022, no 4, https://doi.org/10.3389/fdgth.2022.1010202.

Kaufman G., Kenneth S., *What is systemic risk, and do bank regulators retard or contribute to it?,* „Independent Review" 2003, no 7.

Kyung P., *Data as Public Goods or Private Properties?: A Way Out of Conflict Between Data Protection and Free Speech*, „Irvine Journal of International, Transnational, and Comparative Law" 2021, no 77, https://scholarship.law.uci.edu/ucijil/vol6/iss1/5.

Litwiński P., *Komentarz do art. 4*, [in:] P. Litwiński (red.), *Ogólne rozporządzenie o ochronie danych osobowych. Ustawa o ochronie danych osobowych. Wybrane przepisy sektorowe. Komentarz*, Warszawa 2018.

Maddy O., *Wild and Interesting Facebook Statistics and Facts*, https://kinsta.com/blog/facebook-statistics/.

Melchior C., Oliveira M., *Health-related fake news on social media platforms: A systematic literature review*, „New Media&Society" 2021, no 6, https://journals.sagepub.com/doi/abs/10.1177/14614448211038762.

Muchlinski P., *Multinational Enterprises and the Law*, Blackwell 1999.

Nowak M., *Introduction to the International Human Rights Regime*, The Hague 2003.

Peek N., Holmes J.H., Sun J., *Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics*, „Yearbook Medical Informatics" 2014, no 1, p. 42-47.

Podrecki P., *Komentarz do art. 34*, [in:] M. Grochowski (red.), *Rynek cyfrowy. Akt o usługach cyfrowych. Akt o rynkach cyfrowych. Rozporządzenie platform-to-business. Komentarz*, Warszawa 2024.

Purtova N., *The law of everything. Broad concept of personal data and future of EU data protection law*, „Law, Innovation and Technology" 2018, no 1.

Renn O. *et al.*, *Systemic Risks from Different Perspectives*, „Risk Analysis" 2022, no 42, https://doi.org/10.1111/risa.13657.

Schwemmer C., Knight C., Bello-Pardo E.D., Oklobdzija S., Schoonvelde M., Lockhart J.W., *Diagnosing Gender Bias in Image Recognition Systems*, „Socius" 2020, no 6.

Su D.*et al.*, *Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials*, „Diabetes Research and Clinical Practice" 2016, no 116.

Tar J., *Amazon joins Zalando in challenging very large online platform designation*, https://www.euractiv.com/section/platforms/news/amazon-joins-zalando-in-challenging-very-large-online-platform-designation.

Yanga S., Santillanab M., Koua S.C., A*ccurate estimation of influenza epidemics using Google search data via ARGO*, „Proceedings of the National Academy of Sciences" 2015, no 47.

Dostęp do danych o stanie zdrowia do celów badań naukowych. Uwagi na tle art. 40 Aktu o usługach cyfrowych

Streszczenie

Głównym celem umożliwienia dostępu do danych w oparciu o Akt o usługach cyfrowych jest zwiększenie transparentności działania VLOP i VLOSE, w szczególności poprzez zapewnienie możliwości skutecznej kontroli przez podmioty publiczne oraz zapobieganie negatywnym zjawiskom objętym ryzykami systemowymi poprzez prowadzenie badań przez niezależnych badaczy. Jednocześnie dane, które zbierają VLOP i VLOSE w ramach świadczonych przez siebie usług, bardzo często obejmują informacje o stanie zdrowia swoich użytkowników. Dane te mogą mieć istotne znaczenie w tworzeniu nowych produktów i usług zdrowotnych oraz dla planowania polityk publicznych w tym obszarze. Celem artykułu jest udzielenie odpowiedzi na pytanie, czy dostęp do danych przewidziany w art. 40 Aktu o usługach cyfrowych pozwala na prowadzenie badań w obszarach dotyczących zdrowia. Artykuł podzielony jest na cztery części. W pierwszej przeanalizowano pojęcie danych dotyczących zdrowia, w drugiej zaś rozważano, czy dane medyczne wchodzą w zakres celów umożliwiających udostępnianie danych w oparciu o art. 40 Aktu o usługach cyfrowych. W trzeciej części analizie poddano warunki ograniczające udostepnienie danych do celów badań naukowych. Ostatnia część zawiera wnioski, zgodnie z którymi Akt o usługach cyfrowych pozwala na udostępnianie informacji o stanie zdrowia na potrzeby badań naukowych w tym obszarze, ale jest to bardzo ograniczone przez warunki zawarte w tym akcie. Jednocześnie interpretacja tych warunków może być różnorodna, a od niej w dużej mierze zależy, ile i jakie dane mogą otrzymać naukowcy. Stąd istotną rolę odgrywać będą Koordynatorzy ds. usług cyfrowych, którzy ukształtują praktykę w zakresie realizowania wniosków o dostęp do danych.

Słowa kluczowe

dane o stanie zdrowia, dostęp do danych, VLOP, VLOSE, Akt o usługach cyfrowych, otwarta nauka