Data Ethics and Privacy Essay: Health data are too sensitive to use for anything apart from the original purpose for which they are collected.

David Kennedy

March 2024

1 Introduction

We find ourselves in an age where data is abundant, plentiful, and ever-growing. In 2010, the volume of data created, captured, copied, and consumed worldwide was recorded as being 2 zettabytes (2 trillion gigabytes); in 2025, that number is forecast to surpass 180 zettabytes [22]. While much of this data assessed separately is not of much worth, hidden information, patterns, and trends may be discovered when taken together - this forms the foundational idea of Big Data, which "could be characterized as the value of vast amounts of data, which are of little if any value in small quantities" [18, Section 3].

Naturally, many businesses have been quick to capitalise on this abundance of data, leading to the formation of the 'Big Data analytics' industry: companies which specialise in "examining and analyzing massive and varied data that can help organizations make more-informed business decisions, especially for uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information" [6, Section 1]. As Gandomi et al. mention, "Big data has become essential" to those who must deal with and process massive amounts of information [6, Section 1], and will only continue to become more important as the amount of data increases - this

is reflected in the market value of the industry, which is forecast to rise from 240.56 billion to 655.53 billion from 2021 to 2029 respectively [21].

The Big Data analytics market cannot function in isolation, however - it exists to provide services to other companies and industries. The public sector "is increasingly aware of the potential value to be gained from big data-driven innovation via improvements in effectiveness and efficacy and with new analytical tools" [24, Section 3.2]; the finance and insurance sector use Big Data "to transform their business, realise new revenue opportunities, manage risk, and address customer loyalty" [24, Section 3.3]; and the media and entertainment industries look to Big Data in order to "reduce the costs of operating in an increasingly competitive landscape, and at the same time ... to increase revenue from existing content" [24, Section 3.5].

The healthcare sector also stands to benefit enormously from the introduction of Big Data - the ability to process and analyse large quantities of health data could help to inform developments in a number of areas, from "outcomes and comparative effectiveness research to designing clinical trials and monitoring drug safety" [10, Section 6]. Medical professionals would be able to "analyze large datasets from thousands of patients, identifying clusters and correlation between datasets, as well as developing predictive models" [2, Intro], and therefore may be able to identify medical conditions or diagnose patients far earlier than before as a result.

However, the collection of health data can pose significant risk to the data subject involved, depending on the type of information being stored. "One concern is the possibility that someone could find out something about a person's medical history, and use it against them. Some people are worried about the loss of privacy, damage to their reputation, or discrimination if someone found out about their condition. This may be especially true for anyone with a condition they feel sensitive about." [16, Privacy]

In an attempt to provide the context necessary to address the statement, "Health data are too sensitive to use for anything apart from the original purpose for which they are collected", the following sections clarify both health data and its sensitivity, and then analyse the current regulation of health data reuse from the four perspectives detailed in Lessig's 'Code' [12, Chapter 7]: Law, Market, Norms, and Architecture.

2 Health Data and Sensitivity

Prior to discussing how health data is currently regulated, and using this information to decide whether health data is still too sensitive to reuse, it seems necessary to define precisely what health data is, and how sensitivity is determined. Addressing health data first, it may be most appropriate to use the definition provided in Chapter 1 of the GDPR, as it will serve us well when the legal aspects of health data are covered later. Article 4 of the GDPR defines 'data concerning health' to be "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status" [4, p. 15]. This definition leaves a lot of scope for interpretation - for example, it could be argued that the music genres a data subject frequently listens to could be used to infer information about the subject's mental health and well-being, and therefore that music choice is a weak form of data concerning health. Addressing edge-cases surrounding what is and is not health data is beyond the scope of this paper - rather, the topics discussed and scenarios addressed shall be concerned primarily with data related to health care, as well as the physical health of the data subject.

Another term to clarify is the sensitivity of data. Traditional knowledge would tell us that the more sensitive data is, the greater the risk associated with its misuse or inappropriate disclosure. However, the risks associated with data misuse apply to both data controllers, and data subjects - while the consequences of inappropriate disclosure might involve loss of privacy for the data subject, the consequences for the data controller will look a lot different, ranging from damage to their reputation as a reliable organisation, to legal trouble and fines from a governing body. As is most often done, we will consider the sensitivity of health data from the perspective of the data subject, not the data controller.

As an aside, meaningful statements cannot often be made about the intrinsic sensitivity of health data without first addressing the data subject and the environment they find themselves in. For example, consider a scenario where a patient receives treatment for an ingrown toenail, and details of their treatment are inappropriately disclosed to the public. In and of itself, there isn't anything particularly damaging about the public knowing that the data subject is receiving said treatment, other than perhaps some level of embarrassment. However, perhaps the data subject suffers from severe anxiety, and when they discover that their data has been disclosed to the public, they

lock themselves in their house and refuse to leave ever again, out of fear of ridicule. Clearly, the level of sensitivity of the data subject's treatment is not anything intrinsic to the data itself, but rather it is determined by the data subject. Consider another very similar example, where another patient who does not suffer from anxiety is being treated for an ingrown toenail, and their data is inappropriately disclosed to the public. Here, there should be no issue with extreme reactions from the data subject; however, the subject may happen to live in a society which holds tremendous prejudice against those who suffer with ingrown toenails, and the public band together and exile them forever. While this is admittedly an extreme example, it is useful to convey the fact that the sensitivity of health data cannot be considered apart from the data subject, or their living situation. As Rumbold and Pierscionek discuss, "Socio-cultural factors can dramatically alter the sensitivity of certain data; sexual orientation is much more sensitive in countries where Shariah law is practised, for example. Some characteristics are sensitive only because of the potential for discrimination, for example the categorisation of race and ethnic groups as sensitive in data protection law is due to this. The potential for misuse after accidental or malicious release also affects sensitivity considerations" [18, Section 11].

3 Analysing the current regulation of health data

Having determined what health data is, from which perspective sensitivity is to be considered, and some of the difficulties in determining sensitivity of health data, we can now assess how the reuse of health data has been regulated from the four perspectives outlined in Lessig's 'Code' [12, Chapter 7]: Law, Markets, Social Norms and Ethics, and Architecture.

3.1 The law regarding the use and reuse of health data

Steps have been taken by both the EU and the USA in an attempt to address the risks involved in the reuse of health data - the resulting legal landscape regarding the reuse of health data is incredibly detailed, extensive, and complex. While this results from an effort to provide guidance which covers every likely scenario, it often has the opposite, unintended effect, due to both the legal language used, and the sheer scale of the regulations. Addressing the law in the EU, the European Parliament adopted General Data Protection Regulation in 2016, "one of its greatest achievements

in recent years" [5]. The GDPR "prohibits the processing of health-related personal data (GDPR Art. 9 (1)) unless the informed consent of every affected person is given (GDPR Art. 9 (2a)) or a scientific exemption is present (GDPR Art. 9 (2j))", with the latter being difficult to substantiate [7, Section 3]. From the perspective of the cagey data subject, this regulation is great - unless the data subject gives informed consent for the reuse of their data, it cannot legally be reused. However, from the perspective of the data controller, and researchers operating in the EU in general, this article actually presents a number of problems.

McKeown et al. write that "traditional models of informed consent may be ill suited to big data projects, because these tools were conceived in the context of conventional clinical research such as clinical trials, which are not concerned with the evolving applications and innovative research designs of big data research"; continuing on, "the apparent power of big data analytics derives from its ability to make novel predictive inferences across datasets... this iterative novelty limits what can be communicated to participants about the purposes for which their data may be used" [14, Section 2]. Typical interpretations of informed consent require that the subject must explicitly consent to a specific processing of their data, tying their consent to an existing project; however, Big Data projects are unpredictable by nature, and so oftentimes the researchers may not know precisely what a subject's data will be used for before they conclude the study. This eliminates the possibility of traditional informed consent within the field of Big Data analytics and research powered by AI, as even the researchers themselves oftentimes cannot know what steps or inferences the research model may take.

Considering health data's reuse specifically, think of a scenario where a researcher performs a scientific study with 50 participants, and collects data for one specific purpose, having received consent from each data subject. After the participants leave, the researcher comes up with a fantastic new project similar to the first, which could make use of the data they just collected. Gehrmann et al. mention this scenario as a significant barrier to the progress of the study: "For retrospective projects ... consent cannot be obtained during the patients' stay at the hospital because the project idea does not exist at that time. Hence, the researcher would have to retrospectively contact all patients whose data is needed for the project, describe the project objective and methodology to them and then ask for their consent. This requires great effort, is, itself, questionable in terms of data protection and even not feasible if the patients are deceased." Gehrmann et al. argue that

due to the practical difficulty of carrying out out the process mentioned above, "Making clinical data truly reusable in a research context, therefore, requires a broad consent in which the patients generally agree to the secondary use of their data in ethically approved research contexts" [7, Section 3]. The topic of this broad consent is an ongoing discussion with no single clear solution, with [23] discussing a number of options in the context of consent provided by data subjects to bio-banks.

It is clear that while the regulations currently provided by the law help to protect data subjects from (re)use of their data without their consent, it has also made many aspects of research more difficult, whether it be managing to provide true informed consent in the first place, or having to retrospectively contact data subjects in order to gain their informed consent to facilitate retrospective studies. Any further legal restrictions may run the risk of making health data reuse completely impractical, thereby eliminating a large number of potential studies; the fewer studies there are, the slower the health sector will progress scientifically, and the slower the rate of healthcare improvement will become. This is not simply a slippery-slope argument - since the introduction of the HIPAA privacy rule in 2003, "health researchers have asserted that the Privacy Rule has had a negative effect on researchers' abilities to conduct meaningful research" [15]. While it is incredibly important to protect data subjects from inappropriate disclosure of their information, one must also consider the wide-spread benefits that come from successful medical research and improved public healthcare.

3.2 The state of the health data market

There is significant interplay between the law and the market when it comes to regulating the reuse of health data; not only does the law make it more difficult for companies to reuse health data in order to inform business decisions, but it also may serve as a disincentive for companies that wish to do so. Fears of legal repercussions resulting in financial penalties, reputation damage, and loss of brand value may stifle companies' aspirations to attempt to make use of existing health data repositories without extensive assurance that to do so would not be inappropriate. Therefore, when considering the health data market, we can assume for the purposes of this paper that the companies operating within it will be abiding by the law.

Within any given market, companies compete against each other for a large and loyal customer base; in order to increase the number of customers their business has, they will often offer aspects of the services they provide in different ways than their competitors, in hopes that customers will value these differences. In this way, there is also interplay between social norms and the market; depending on what society values, companies that manage to provide for these values will be rewarded with growth accordingly. For example, two competing healthcare companies might offer different data privacy schemes - the company which offers the scheme which most aligns with customers' preferences will likely receive more business. However, a choice is to be made between providing complete privacy in hopes of attracting more customers, and increasing sales and revenue by selling customer data; depending on the size of the company and its values, choices may vary.

Trust also plays a large part in the success of a company within the health data market, as will be reinforced in the following section - if the data subject trusts the organisation that wished to reuse their data, they will be more willing to provide their consent, and the organisation will flourish as a result. This aspect of the relationship between company and customer does introduce the risk of ethics washing - companies will attempt to appear to be more ethical than they actually are, in hopes that this will increase the level of trust customers are willing to give them, and in turn, increase their profits.

For the most part, in the case of health data reuse, the market is largely subject to the law of the land, and the social norms of the customers - it would be difficult for a company to effect any meaningful change regarding reuse practices, other than perhaps lobbying for legal change.

3.3 Social norms and ethics regarding the reuse of health data

Along with the law, social norms and ethics effect the largest regulatory power on the reuse of health data; norms and ethics influence legal standards (and vice versa), both of which influence changes in the market. In 'A review of attitudes towards the reuse of health data among people in the EU' [19], researchers highlight a clear split of attitudes towards the reuse of health data, one which is somewhat unsurprising. Data subjects generally had positive opinions towards the reuse

of their health data, provided they had the perception that the reuse will serve the common good (3.2); conversely, data subjects saw the reuse of health data as unacceptable when it either was not perceived to serve the common good, or was seen as conflicting with their interests [19, Section 3.3].

This split in attitudes is a perfect example of society conforming to the ethical framework of utilitarianism [3] - while the individual data subject runs the risk of having their health data misused, they perceive the potential for positive change as a result of their data reuse as more valuable than their potential suffering. Granted, this might not be true utilitarianism, as individuals may be additionally selfishly motivated by their identification of themselves within those who may benefit from the common good, but the results are the same regardless. Using the same ethical framework, weighing the potential prosperity of a company which commercialises health data against the potential of having health data misused results in a tilting of opinion in the opposite direction - individuals are likely to care very little for the success of companies in comparison to the common good, as they cannot identify themselves with the given company.

A similar study was run in Scotland to gauge public sentiment towards engaging in the Scottish Health Informatics Program [1, Section 4], with very similar findings - participants were happy to share their data if they perceived it to be for the common good, and uncomfortable with doing so in the cases where private companies would be profiting. However, it is important to acknowledge the roles that perceived informed consent and trust play in this process - we need only look to the controversy of the 'care.data' scheme [20]. Despite the NHS describing the development of the care.data program as being done with the aim to "ensure that the best possible evidence is available to improve the quality of care for all" [20, Section 2], the public were incredibly unhappy with the way the scheme had been implemented.

Some of the public's misgivings included lack of transparency, lack of respect for confidentiality and privacy, misgivings about the opt-out scheme, and worries around commercialisation [20, Section 3] - the UK public did not trust the government with centralised access to their health data. It seems clear that society's standard for the reuse of health data demands that the reuse will serve the common good, rather than benefit an individual organisation or private company; moreover, the data subject must trust the future data processor, and should be treated with respect throughout the process. If this standard is not met, public outrage will ensue.

3.4 Technologies and practices used in the reuse of health data

There are several practices used in the health data industry, informed by the law and the available technology, which are used to attempt to secure data subjects' information, and maintain their right to privacy and confidentiality. Arguably the most important practice is storing data securely, for which many technologies have been created. According to Kruse et al.[11, Results], the two most frequently discussed security techniques when it comes to health data are firewalls and cryptography; these two techniques address the security of the entry points into a system, and the protected payload of a system, respectively. A firewall "is a network device that enforces security policy for network traffic" - it creates a barrier between separate networks, acting as a screening point used to reject unwanted traffic [17]. In the case where a hacker attempts to access a network in order to steal valuable health data, a firewall would ideally prevent the hacker from joining the network altogether, ensuring that they don't get access to the data. Data encryption, on the other hand, is "the process of transforming data to make it unreadable except to those possessing some secret knowledge, usually referred to as a key" [13, Definition]; in the previous case, if the firewall failed and the hacker joined the network, they wouldn't be able to make sense of the health data if it were encrypted, and as such, would be worthless to them without the key. Encryption isn't only used for data at rest - "specifically, encryption has enhanced security of EHRs during the exchange of health information" [11, Results]. Encrypting health data before transmission ensures its protection against meaningful interception by eavesdroppers and attackers, as they could not make sense of the data without the key. Securing the port of access to a data repository, as well as the data itself, falls within good practices when dealing with health data, and within good security practices in general.

Another practice commonly used in the health data industry is anonymisation of data. While GDPR requires that you do not keep data for any longer than you need it for, it allows for the retention of that data in anonymised form; this is any data which is "no longer in a form which permits identification of data subjects" [9]. Retention of data in an anonymised form allows for potential secondary use in the future, while respecting the privacy and confidentiality of the original data subjects, and eliminates any risks associated with data leaks, as theoretically, none of the data could be linked back to the people who provided the data. However, it is important to recognise that this is a constantly evolving field, and there are sometimes cases where data can be linked back to the original data subject through inference techniques [8].

4 Addressing the statement, "Health data are too sensitive to use for anything apart from the original purpose for which they are collected."

To say 'Health data are too sensitive' is to make a blanket statement that doesn't do the complexity and variability of the subject matter justice. As discussed previously, the sensitivity of health data depends not only on its content, but also on the data subject and their geo-religio-political environment; as such, the sensitivity of health data can only be evaluated accurately on a case-by-case basis. Nevertheless, it is impossible to ignore the fact that sensitive health data certainly exist, and deserve to be protected as best as possible. For this reason, among others, the decision has been made to treat all health data as sensitive, and mandate the appropriate protective steps and restrictions by law for dealing with them.

> Specific informed consent must be provided before health data is collected, processed, or (re)used, and data must either be anonymised or deleted once it has been used for its original and intended purpose. While stored, both the data and the system responsible for the data should be secured as a matter of professional practice; if the expected standards are not maintained, there will be public unrest, and the organisation will likely suffer both legal and financial penalties as a result. Additionally, individuals will only consent to having their data processed if they evaluate it to be for the common good, not contrary to their own interests, and if they trust both the organisation's intentions, and the organisation itself.

The rule-set laid out above is the system inside which the majority of society find it reasonable for organisations to collect, process, and use health data. Outside of these rules, health data are perhaps too sensitive to use for anything at all. However, if data controllers and processors abide by the regulations discussed throughout this paper, the majority of people believe that they are given sufficient agency to weigh up the potential benefits and drawbacks of passing on their data, and to make the decision for themselves. When the time comes that the majority decide new changes must be made regarding the regulation of health data collection and processing, new legal, ethical, practical, and technological developments will be produced to satisfy these standards.

References

[1] Mhairi Aitken, Sarah Cunningham-Burley, and Claudia Pagliari. "Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme". In: *Science & Public Policy* 43.5 (Oct. 2016), pp. 713–723. ISSN: 0302-3427. DOI: 10.1093/scipol/scv075. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210028/ (visited on 03/07/2024).

[2] Kornelia Batko and Andrzej Ślezak. "The use of Big Data Analytics in healthcare". In: *Journal of Big Data* 9.1 (2022), p. 3. ISSN: 2196-1115. DOI: 10.1186/s40537-021-00553-4. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8733917/ (visited on 03/07/2024).

[3] Julia Driver. "The History of Utilitarianism". In: (Mar. 27, 2009). URL: https://seop.illc.uva.nl/entries/utilitarianism-history/ (visited on 03/07/2024).

[4] EU. *Art. 4 GDPR – Definitions*. General Data Protection Regulation (GDPR). 2018. URL: https://gdpr-info.eu/art-4-gdpr/ (visited on 03/07/2024).

[5] EUROPEAN DATA PROTECTION SUPERVISOR. *The History of the General Data Protection Regulation — European Data Protection Supervisor*. May 25, 2018. URL: https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en (visited on 03/07/2024).

[6] Amir H. Gandomi, Fang Chen, and Laith Abualigah. "Machine Learning Technologies for Big Data Analytics". In: *Electronics* 11.3 (Jan. 2022), p. 421. ISSN: 2079-9292. DOI: 10.3390/electronics11030421. URL: https://www.mdpi.com/2079-9292/11/3/421 (visited on 03/07/2024).

[7] Julia Gehrmann et al. "What prevents us from reusing medical real-world data in research". In: *Scientific Data* 10.1 (July 13, 2023), p. 459. ISSN: 2052-4463. DOI: 10.1038/s41597-023-02361-2. URL: https://www.nature.com/articles/s41597-023-02361-2 (visited on 03/07/2024).

[8] Melissa Gymrek et al. "Identifying Personal Genomes by Surname Inference". In: *Science* 339.6117 (Jan. 18, 2013), pp. 321–324. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1229566. URL: https://www.science.org/doi/10.1126/science.1229566 (visited on 03/07/2024).

[9] Information Commissioner's Office. *Principle (e): Storage limitation*. Principle (e): Storage limitation. Aug. 4, 2023. URL: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/the-principles/storage-limitation/ (visited on 03/07/2024).

[10] Bonnie Kaplan. "How Should Health Data Be Used?: Privacy, Secondary Use, and Big Data Sales". In: *Cambridge Quarterly of Healthcare Ethics* 25.2 (Apr. 2016), pp. 312–329. ISSN: 0963-1801, 1469-2147. DOI: 10.1017/S0963180115000614. URL: https://www.cambridge.org/core/journals/cambridge-quarterly-of-healthcare-ethics/article/abs/how-should-health-data-be-used/D3762C502A4C38EA79F1B516AD3665D2 (visited on 03/07/2024).

[11] Clemens Scott Kruse et al. "Security Techniques for the Electronic Health Records". In: *Journal of Medical Systems* 41.8 (July 21, 2017), p. 127. ISSN: 1573-689X. DOI: 10.1007/s10916-017-0778-4. URL: https://doi.org/10.1007/s10916-017-0778-4 (visited on 03/07/2024).

[12] Lawrence Lessig and Lawrence Lessig. *Code*. Version 2.0. OCLC: ocm77638613. New York: Basic Books, 2006. 410 pp. ISBN: 9780465039142. URL: https://lessig.org/product/codev2.

[13] Ninghui Li. "Data Encryption". In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 574–574. ISBN: 9780387399409. DOI: 10.1007/978-0-387-39940-9_98. URL: https://doi.org/10.1007/978-0-387-39940-9_98 (visited on 03/07/2024).

[14] Alex McKeown et al. "Ethical Issues in Consent for the Reuse of Data in Health Data Platforms". In: *Science and Engineering Ethics* 27.1 (Feb. 4, 2021), p. 9. ISSN: 1471-5546. DOI: 10.1007/s11948-021-00282-0. URL: https://doi.org/10.1007/s11948-021-00282-0 (visited on 03/07/2024).

[15] Sharyl J. Nass et al. "Effect of the HIPAA Privacy Rule on Health Research". In: *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press (US), 2009. URL: https://www.ncbi.nlm.nih.gov/books/NBK9584/ (visited on 03/07/2024).

[16] NHS Confederation. *What are the risks around patient data? — Understanding patient data*. What are the risks around patient data? 2024. URL: https://understandingpatientdata.org.uk/weighing-up-risks (visited on 03/07/2024).

[17] Niels Provos. “Firewall”. In: *Encyclopedia of Cryptography and Security*. Ed. by Henk C. A. van Tilborg. Boston, MA: Springer US, 2005, pp. 230–233. ISBN: 9780387234830. DOI: 10.1007/0-387-23483-7_169. URL: https://doi.org/10.1007/0-387-23483-7_169 (visited on 03/07/2024).

[18] John M. M. Rumbold and Barbara K. Pierscionek. “What Are Data? A Categorization of the Data Sensitivity Spectrum”. In: *Big Data Research*. Big Data Centric Computational Intelligence in Bioinformatics and Healthcare 12 (July 1, 2018), pp. 49–59. ISSN: 2214-5796. DOI: 10.1016/j.bdr.2017.11.001. URL: https://www.sciencedirect.com/science/article/pii/S2214579617302010 (visited on 03/07/2024).

[19] Lea L. Skovgaard, Sarah Wadmann, and Klaus Hoeyer. “A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good”. In: *Health Policy (Amsterdam, Netherlands)* 123.6 (June 2019), pp. 564–571. ISSN: 0168-8510. DOI: 10.1016/j.healthpol.2019.03.012. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6558994/ (visited on 03/07/2024).

[20] Sigrid Sterckx et al. ““You hoped we would sleep walk into accepting the collection of our data”: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research”. In: *Medicine, Health Care, and Philosophy* 19 (2016), pp. 177–190. ISSN: 1386-7423. DOI: 10.1007/s11019-015-9661-6. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4880636/ (visited on 03/07/2024).

[21] Petroc Taylor. *Big data analytics market size worldwide 2029*. Statista. Feb. 21, 2024. URL: https://www.statista.com/statistics/1336002/big-data-analytics-market-size/ (visited on 03/07/2024).

[22] Petroc Taylor. *Data growth worldwide 2010-2025*. Statista. Nov. 16, 2023. URL: https://www.statista.com/statistics/871513/worldwide-data-created/ (visited on 03/07/2024).

[23] Rachel Thompson and Michael J. McNamee. “Consent, ethics and genetic biobanks: the case of the Athlome project”. In: *BMC Genomics* 18 (Suppl 8 Nov. 14, 2017), p. 830. ISSN: 1471-2164. DOI: 10.1186/s12864-017-4189-1. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5688492/ (visited on 03/07/2024).

[24] Sonja Zillner et al. “Big Data-Driven Innovation in Industrial Sectors”. In: *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. Ed. by

José María Cavanillas, Edward Curry, and Wolfgang Wahlster. Cham: Springer International Publishing, 2016, pp. 169–178. ISBN: 9783319215693. DOI: 10.1007/978-3-319-21569-3_9. URL: https://doi.org/10.1007/978-3-319-21569-3_9 (visited on 03/07/2024).