Big Biological Data: Need for a Reorientation of the Governance Framework

Esra Demir, LLM
PhD Candidate at Erasmus School of Law
Junior Fellow at Jean Monnet Centre of Excellence on Digital Governance
Rotterdam, The Netherlands
demir@law.eur.nl

2022 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) | 978-1-6654-8462-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/CIBCB55180.2022.9863047

***Abstract*— Biological data (biodata) has become a buzzword for innovations in this century. With pioneering advances in machine learning techniques and artificial intelligence, biological datasets have brought about revolutionary changes in many areas, particularly in healthcare. In this process, a host of ethical and legal issues have also come to the fore. The purpose of this paper is to address the need for a reorientation of the governance framework for human biodata. Using the case of Dinerstein v. Google, this paper argues that the data governance framework must take into account and balance interests of both fostering innovation and protecting societal values.**

***Keywords*— big data, biodata, data governance, social values, innovation**

I. Introduction

With digitization and datafication, masses of information with and about human bodies and their daily lives have been generated, collected, and stored [1]. On the one hand, collections of biodata play an essential role, especially in biotechnological innovations and biomedical research; therefore, the use of biodata is of great importance. On the other hand, numerous legal and ethical questions have arisen in society concerning privacy, autonomy, democracy and so on [2]. The purpose of this paper is to indicate the need for a reorientation of the current data governance framework for human biodata, arguing that the current approach to data governance is insufficient to address the legal and ethical complexities. Therefore, data governance should be reoriented, taking into account and balancing the interests of both promoting innovation and protecting societal values.

This paper, first, presents the concept of big biological data. This section addresses the increasing amount of biodata and highlights the variety and value of these datasets in different areas, especially for technological innovations and research. Second, it outlines the challenges of data governance from a legal perspective. To illustrate the complexity of how biodata can be governed to both promote innovation and protect societal values, this section presents and discusses a court case, the Dinerstein v. Google case. Third, using the information provided in the previous sections, it examines why the governance framework needs to be reoriented and how this can be done. Although there are many valuable types of biodata derived from studies of animals, plants, microbial organisms, and populations, this paper limits itself by focusing primarily on human biodata.

II. Big Biological Data

A. How Do Big Biological Data Come To Matter?

Historically, data have always been important; however, with computerization, the world has been flooded with more data than ever before [3]. Big data as a term was added to the literature in the late 20th century by sciences such as astronomy and genomics, which were the first to experience the data explosion [3]. As with many other terms used to describe the rapidly growing use of technology and processes, there is no precise and widely agreed upon definition of big data [3,4]. It initially refers to the volume of data that cannot be efficiently processed by using traditional database methods and tools as a result of the complexity and high level of variability [5,6].

The rapid development and innovation of high-throughput sequencing technologies have enabled biology to enter the big data era; consequently, more biologists have become involved in big data [7,8]. In this regard, biologists started to generate an enormous amount of sequence data, and computers as data processors provided a ready-made instrument for biodata [9]. Therefore, biology has become a computerized and computerizable discipline [9]. The combination of technological developments with the increase in biological knowledge has made the acquisition of big biological data possible [8]. And the concepts of 'biology – data – information' have come into use together more than ever [10].

In recent years, for various reasons such as treatment, ancestry testing, and research purposes more and more biological materials and data are extracted from individuals and stored in biobanks, and the obtained information are digitally recorded [11]. With the myriad number of high-throughput sequencing projects, the amount of available biodata is still increasing [12]. According to a study in 2015, the total amount of sequence data produced has been doubling about once every seven months. And if the growth keeps continuing at the current rate, it is predicted to reach one zetta-base pairs of sequence per year by 2025 [13].

B. What Does This Massive Amount of Data Entail?

Biological datasets are highly complex and fragmented compared to other datasets [14]. Essentially, it can be said that there are two types of biodata: derivative data and descriptive data [15]. The former refers to data derived directly from organisms and individuals, such as tumor tissue, cells, blood, DNA, and DNA microarray results. The latter refers to data used to describe the biology of individuals and their lifestyles, such as their response to the environment, experience of disease, risk of mortality, and social identity [15]. This dual structure shows that biodata may exist in the material form such as DNA embedded in tissues, which is different from classical data understanding. Therefore, biodata might be sequences, graphs, geometric information, scalar and vector fields, patterns, constraints, images, and spatial information [16].

Biological datasets are not only vast and various, but also contain perhaps the most valuable and sensitive information about individuals. They can perform different functions in different areas, such as a source of forensic evidence, a source

978-1-6654-8462-6/22/$31.00 ©2022 IEEE

of information for transplantation, a source of therapeutics, and a source of diagnostics. These datasets are indeed of vital importance for biotechnological innovations and biomedical research. They offer numerous opportunities, particularly for predictive, preventive, personalized and participatory medicine [8]. To reveal the significant information hidden in these datasets, artificial intelligence and machine learning techniques provide useful tools. By using artificial intelligence to create diagnostic devices and equipment, service providers collect biodata, upload them to the cloud or a centralized data center to analyze or diagnose, and then receive instructions based on the individual's specific needs [17]. One of the groundbreaking results is the improvement of deep learning algorithms that can detect skin cancer lesions from images, as dermatologists do [18].

Not surprisingly, the accumulation of large amounts of these valuable datasets necessitates revolutionary measures for data management, analysis, and accessibility [8,19]. These needs of big biological data can be illustrated as large storage capacity, real-time analysis, and secure integration of distributed datasets [20]. In this regard, the importance of developing data mining and analysis techniques, efficient, sensitive, and better able to handle big biological data, has been also increased [21]. For efficient, scalable, and productive analysis of biodata, researchers, especially engineers and data scientists, have made great efforts to invest in developing proper tools [3,12]. Apart from the technical issues, there are, however, numerous legal and ethical questions concerning the big biological datasets. The next section discusses the challenges of data governance from a legal perspective.

III. Challenges for the Governance of Human Biodata

In today's data environment, human biodata can be generated by everyone, not just professionals, and by everything including public and private data aggregation initiatives, biobank collections, wearable devices, and so on [5,22]. With the increase use of digital technologies in daily routine, it has become possible to expose, produce, isolate, aggregate, process, analyze, buy and sell, exploit, transfer, and circulate them [23]. It is also possible to apply these aggregated datasets to algorithms for the purpose of predicting future behavior and making judgments about individuals and groups [2]. In this context, concerns regarding societal values such as privacy, autonomy, justice, equality, democracy, etc. have become more prevalent. Therefore, questions related to data governance, such as by whom and for what purposes can human biodata be used and how can the data be traced by individuals, clearly need to be answered [20].

To illustrate the legal complexities in data governance, the class action lawsuit against Google and the University of Chicago will be presented [24]. It should be noted that although the US case is used, similar cases can also be found all over the world. For example, the collaboration between NHS and Google DeepMind is another case that occurred in the UK [25, 26]. Here, the Dinerstein v. Google case will be used to demonstrate patient privacy concerns in the context of data sharing and the use of AI.

A. Dinerstein v. Google Case

On May 17, 2017, the University of Chicago (UChicago) Medicine announced a research partnership with Google to explore ways to use data from electronic health records (EHRs) to build technology that could improve the quality of healthcare services [27,28]. With this partnership, they have focused on developing predictive health models to reduce unexpected hospital readmissions and anticipate future medical events by using new machine-learning techniques [27]. As part of this research partnership, the UChicago disclosed/transferred to Google EHR data of all adult patients treated at its hospital from January 1, 2010, until June 30, 2016 [24].

The results of this research were published as an article in 2018, namely 'scalable and accurate deep learning with electronic health records' [29]. With this study, deep learning models achieved high accuracy for tasks such as predicting in-hospital mortality, unplanned 30-day readmissions, long length of stay, and discharge diagnosis [29]. As pointed out in the methodology part of the article, all EHRs were de-identified; however, the dataset provided by the UChicago Medicine contained dates of services and free-text medical notes [29]. This additional information in UChicago Medicine's dataset prompted a class action complaint filed on June 26, 2019, by the law firm Edelson PC on behalf of Matt Dinerstein and all other patients in the same situation [30].

Plaintiff Matt Dinerstein was an inpatient, hospitalized twice, at UChicago Medicine in June 2015. He brought multiple claims against the UChicago, the UChicago Medical Centre, and Google on behalf of himself and all patients whose EHRs disclosed [24]. In the complaint, which seeks monetary damages and an injunction to stop the use and further transfer of patient records, he alleged the violation of state consumer protection laws, breach of (express and implied) contract, violation of common law privacy rights, as well as other claims [24,31]. In general, the plaintiff argued that the Healthcare Insurance Portability and Accountability Act (HIPAA) had been violated by both Google and UChicago by compromising patient privacy. The plaintiff claimed that the additional information in the UChicago dataset - dates of treatments and free-text notes - would be 'a prima facie violation of HIPAA' [24]. According to the plaintiff, although the forms he signed at the hospital specified that the patient's identity and the identity of his medical records will not be included in any research findings or reports, UChicago transferred to Google a large number of patients' EHRs that had not been properly anonymized; therefore, patients' privacy was at risk [24]. Furthermore, he argued that the risk of re-identification was significant because, as a major tech giant, Google holds a great deal of information about individuals through its other services. With the combination of the information collected, Google can identify who the patients in those records are [24]. The plaintiff also alleged that UChicago did not obtain patients' express consent before sharing their EHRs with Google pursuing commercial purposes [24].

On September 4, 2020, the Illinois District Court dismissed all claims brought by the plaintiff against Google and the UChicago [24]. The judge of District Court of Illinois, Rebecca Pallmeyer, ruled, among other things, that plaintiff Dinerstein had not shown that the university had financially harmed him by sharing EHR data with Google [24,32,33]. This means that, in order to claim damages, patients must show that the value of their medical records has been diminished as a result of the invasion of privacy [33]. Dinerstein appealed the decision [32].

B. The Challenges in the Governance of Human Biodata

The case of Dinerstein v. Google has the potential to become a pioneering case regarding when and under what circumstances data might be shared, whether data can truly be anonymized, and what possible safeguards can be in place [34]. While not the only case to raise these issues, this case does provide a good example to demonstrate the complexity of human biodata governance.

On the one hand, it shows how difficult technology companies are to deal with data when they enter one of the most promising areas of artificial intelligence - *the diagnosis of medical problems* [35]. Technology firms, from start-ups to behemoths, are eager to gain access these datasets in order to develop products for predictable next-generation healthcare [31]. As is well known, Google is at the forefront of an effort to improve the technology for reading electronic medical records and help clinicians achieve high diagnostic accuracy in the analysis of medical conditions [35]. To achieve this goal, a huge amount of biodata collected by hospitals, technological devices and other institutions need to be analyzed by machines to learn the skills [35].

On the other hand, in terms of data sharing and use, it covers a number of salient and unresolved questions about uncertainty, concerns about data security, and whether data collected by institutions and devices are properly protected [22,34]. Concerns about legal and ethical challenges may arise from consequentialist concerns or from deontological concerns [36]. Consequentialist concerns arise from negative consequences, which may include tangible negative consequences such as discrimination, extra costs, or stigma, as well as the emotional distress of knowing that such data is public and can potentially be exploited [36]. Deontological concerns do not arise from experiencing negative consequences, but from an invasion of privacy that manifests itself even if no one uses it or if the person never realizes that an invasion has taken place. Here the intrusion is an ethically problematic issue rather than damage in the consequentialist sense [36]. From a legal point of view, these concerns must be addressed, and the protection of social values must be ensured in the data governance framework.

Admittedly, the use of AI and big data, particularly in healthcare, has enormous potential to change healthcare for the better. However, as some reflected with the Dinnerstein v. Google case, it comes with many ethical and legal challenges, such as a lack of legal certainty, trust, transparency, liability, informed consent, algorithmic fairness and bias, data protection and privacy, and cyber security [37]. In the collaboration between the NHS and Google DeepMind, for example, Google DeepMind was given access to a wide range of data on its 1.6 million patients to create a smartphone app called Streams for clinical teams to combat acute kidney injury through the information-sharing agreement [25,38]. However, this cooperation was found problematic because it was not clear what kind of data the transfer contained - millions of identifiable personal medical records [25]. In 2017, the UK Information Commissioner's Office (ICO) ruled that the UK Data Protection Act had been breached since individuals had not been properly informed of the processing of their data as part of the research [39]. Furthermore, it was found that the use of Streams has not had a statistically significant beneficial effect on the clinical outcome of patients [40].

The growing promise of AI to improve healthcare through machine learning systems and existing concerns about and distrust of powerful, lightly regulated, large tech companies, demonstrates the need for a *thoughtful approach* to data governance [31]. The next section examines the need for a reorientation of the data governance framework by looking at the current state of the art in the US and the EU and how this can be achieved.

IV. Necessity of Reorienting the Governance Approach

Given the *sui generis* nature of biodata - the dual structure consists of descriptive and derivative forms - it would not be wrong to say that human biodata plays an essential role in who we really are. And human biological datasets, which may contain the most valuable and sensitive information, are an important source for biotechnology innovation and biomedical research. However, the current data environment, consisting of digitization and datafication, not only erodes our privacy, but also touches on the fundamental conditions of being human [41]. As illustrated in Fig. 1, an approach to data governance should therefore aim at protecting individuals and addressing societal values, while ensuring the reliable availability of data and promoting innovation.

A. The Current Data Governance Approaches

Current governance approach through data regulations have been criticized for failing to adequately address raising legal and ethical problems. In the US, data regulation is sector specific. HIPAA provides protection for patient health information [42]. It has long been debated that HIPAA, which

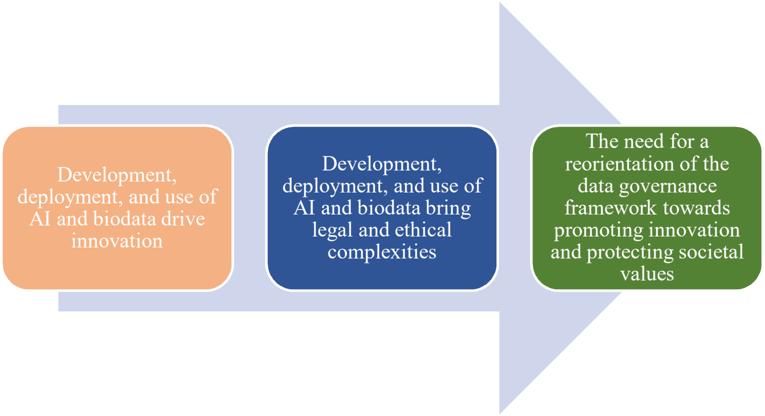
Development, deployment, and use of AI and biodata drive innovation
Development, deployment, and use of AI and biodata bring legal and ethical complexities
The need for a reorientation of the data governance framework towards promoting innovation and protecting societal values

Fig. 1. Towards a thoughtful approach in the biodata governance framework

came into effect in 1996, is already outdated and ill-equipped for data practices in the 21st century. The Dinerstein case also confirms this reality. It means that the traditional remedies provided by this law are inadequate to identify and solve the current challenges posed by digitization and datafication [31]. In Europe, there have been numerous attempts and legislative activities, both at member state and union level, to address data governance issues. It can be said that a holistic approach to data governance has been adopted in Europe. The General Data Protection Regulation (GDPR), which came into force in 2018, regulates personal data protection in general [42]. By comparison, the scope of the GDPR is broader than that of HIPAA, which only provides protection for specific types of health information produced by covered entities and their business partners [37]. Although the GDPR is a relatively new and promising regulation, it has also been criticized for not sufficiently addressing data governance issues [43]. For example, potential harm resulting from the use of data often falls outside the scope of governance frameworks that use traditional legal concepts [44]. Even if the harm is the result of a breach of the law, it may be impossible to demonstrate a causal link between the use of the data and the harm for the purposes of a claim for compensation under the GDPR [43]. This shows that when Dinerstein-like cases arise, individuals may struggle to find legal protection.

As a matter of fact, both in the US and in the EU, activities in the field of data protection regulation are continuing. California, for example, has taken action at the state level and adopted the California Consumer Privacy Act in 2018, which came into effect in 2020. This law provides protection for health data collected outside the health system - which are not covered by HIPAA [37]. More recently, on May 4, 2022, the European Parliament and the Council adopted a new law - the Data Governance Act, which will apply 15 months after the entry into force of the Regulation [45]. This new law may provide a solution to some of the shortcomings and drawbacks of the current regulatory framework that hinder the sharing and reuse of data for biomedical research [46].

Here it must be noted that the problems regarding data governance are multifaceted. It has been said that the current approach to data governance is based on individualism [47,48]. In other words, the laws address the effects of data at the individual level, not at the population level. And, in the current system, the legal relationship is grounded in notice and consent [49]. The practice of notification and consent - people usually do not know what exactly they are consenting to - shows that this has become a symbol of compliance [50]. This approach therefore misrepresents the problems and is incapable of solving them. It does not identify what effects can cause harm and how these effects can be constructed to generate shared benefits. This is because individualistic conceptions of information harm fail to show the social effects of data production [49]. Given the multifaceted nature of the problem, the solutions regarding data governance must also be multifaceted.

B. How Can the Reorientation Be Achieved?

At the global level, strategies for the ethical and legal debate on AI and big data and their applications in healthcare have been discussed, and guidelines for the regulation of AI applications have been developed. For example, in January 2020, the White House published draft guidelines for the regulation of AI applications, which contain 10 principles that agencies should consider when formulating approaches to AI applications: public trust in AI; public participation; scientific integrity and information quality; risk assessment and management; benefits and costs; flexibility; fairness and non-discrimination; disclosure and transparency; safety and security; and interagency coordination [51]. Followingly, in February 2022, the White House Office of Science and Technology Policy (OSTP) clarified their mission on AI as maximizing the benefits of science and technology to promote health, prosperity, safety, environmental quality, and justice for all Americans [52]. Meanwhile, in 2019, the High-Level Expert Group on Artificial Intelligence (AI HLEG) presented the ethical guidelines for trustworthy AI to the European Commission to help promote responsible and sustainable AI innovation. On July 17, 2020, the AI HLEG published their final assessment list for trustworthy artificial intelligence [53]. According to this, the concept of trustworthy AI presents seven main requirements, which are human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability [53]. By making ethics a core pillar of AI, this approach aims to benefit, strengthen, and protect both individual human flourishing and the overall well-being of society [53].

Indeed, the trustworthy AI vision has emphasized that adequate data governance is necessary to prevent damage to societal values. Such data governance helps to self-assess the compliance of the AI system and datasets [53]. In this context, data governance must address the quality and integrity of the data/datasets used, their relevancy considering the domain in which the AI systems will be deployed, and the access protocols and the ability of processing data in a way that protects societal values [53]. Indeed, the quality of datasets directly affects the performance of AI systems. When data are collected, they may contain socially constructed biases, inaccuracies, errors, and mistakes [54]. The governance approach should take this issue into account before training with certain datasets. Additionally, the integrity of the datasets must also be ensured. The insertion of malicious data into an AI system can change its behavior, especially in the case of self-learning systems. The processes and datasets used must be tested and documented at every step, such as planning, training, testing, and deployment [37,54]. For example, biological datasets used by AI systems may suffer from the inclusion of unintentional historical bias, incompleteness, and a poor governance model. The persistence of such biases could lead to direct or indirect prejudice and discrimination against certain groups. It is therefore important to eliminate identifiable and discriminatory biases at the data collection stage wherever possible [53].

The governance mechanism must provide protection throughout the system's life cycle. This means that the framework must not only provide protection for data initially provided by individuals, but also for data generated about individuals throughout their interaction with the system [54]. Therefore, the data governance framework should be enhanced by using the technical and non-technical methods. These methods can be variable, such as systems architecture, design-based solutions, regulation, standardization, etc. [54]. Here it is worth mentioning the design-based ideas, which have become increasingly prominent in solution mechanisms in recent years, and are also widely applied, such as Article 25 of the GDPR - data protection by design and by default. Design-based solutions establish precise and explicit links between the abstract principles and values to be observed by the governance system and the specific implementation actions. At the heart of this method is the idea that compliance with norms can be implemented in the design of the AI system. Design-based solutions can help ensure public trust in AI systems and big data by securing the processes, data, and outcomes [54].

To achieve a desirable societal goal of AI benefiting everyone, the data governance framework requires public trust [37]. In order for individuals to trust the data collection process, it is important to ensure that the collection of data about them will not be used to discriminate against them unlawfully or unfairly [54]. Therefore, the use of data or datasets must be rigorously assessed, especially for new AI-based technologies, which will be deployed in the clinical domain. For example, Article 35 of the GDPR mandates a data protection impact assessment of the intended processing activities for high-risk data processing activities [37,55]. In addition to that, the data governance framework also requires data protocols that regulate access to the data. With these protocols, it is clarified who has access to the data and under what circumstances. Certain restrictions on data access, such as allowing only qualified individuals to access the data, will strengthen trust in the data governance system [54].

To build a relationship of trust, consultation with stakeholders who are directly or indirectly affected by the system is advisable [53]. It is crucial to collaborate with stakeholders to see all aspects and address the ethical and legal concerns in AI and big data, especially in healthcare. In the governance of human biodata, interests need to be harmonized and balanced. Parties who have interests in biodata are quite diverse including individuals, researchers, research institutes, pharmaceutical companies, public and private initiatives, AI makers, and governmental bodies, etc. The interests of these parties are also diverse. They may stem from purely monetary desires, concerns about discriminatory acts or stigmatization, or scientific aspirations to live in a better society, better healthcare. These interests must be taken into account and balanced at both group and individual levels.

The governance framework must be adapted to strengthen not only individual but also collective control and accountability in the context of data use [43]. As human beings, we must be aware of the value of our biodata, but we must also be aware of the importance of pursuing biotechnological innovations and biomedical research. The governance approach is aimed not only at ensuring appropriate protection of individuals and groups, but also at making data available in a reliable manner. With the awareness of the importance of human biodata in the scientific, commercial, and social fields, the collective responsibility of the parties is essential for the reorientation of the governance framework [43]. To this end, it is also important to have public and political discussions on ethical and legal issues, such as the impact of an AI-driven healthcare system on society, and to reconsider and update the existing governance framework in light of new technological developments [37].

In conclusion, this paper attempted to address the need for a reorientation in the field of human biodata governance. To this end, background information on human biodata was first presented by addressing the increasing amount of biodata and highlighting the variety and value of these datasets. Then, using a court case, Dinerstein v. Google, the legal and ethical challenges related to data governance were illustrated. Finally, given the current governance approach in the US and the EU, the need for a reorientation of the governance framework was emphasized and some possible ways of reorientation were explored to balance the protection of societal values and the promotion of innovation. In this context, the data governance framework should, for example, pay attention to data quality and integrity and take it into account at every possible stage in order to minimize ethical and legal concerns. Furthermore, the data governance framework should provide protection through technical and non-technical measures throughout the life cycle of the system. Moreover, to achieve a desirable societal goal where AI benefits everyone, public trust must be ensured. Therefore, cooperation between all parties with an interest in biodata must be improved. For this purpose, collective responsibility is vital.

Acknowledgment

E. D. thanks the anonymous reviewers, Prof. Dr. Klaus Heine, Prof. Dr. Evert Stamhuis, and Dr. Alberto Quintavalla for their valuable feedback. Thanks are also gratefully acknowledged to Erasmus Initiative Societal Impact of AI (AiPact) and Erasmus School of Law for their financial support. The usual disclaimer applies.

References

[1] D. Lupton, "How do data come to matter? Living and becoming with personal data," Big Data & Society, pp. 1-11, July 2018.

[2] J. S. Winter and E. Davidson, "Big data governance of personal health information and challenges to contextual integrity," The Information Society, vol. 35(1), pp. 36-51, 2019.

[3] V. Mayer-Schönberger and K. Cukier, Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013.

[4] R. Kitchin, "Big data and human geography: Opportunities, challenges and risks," Dialoguesin Human Geography, vol. 3(3), pp. 262-267, December 2013.

[5] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," 46th Hawaii International Conference on System Sciences, IEEE, pp. 995-1004, March 2013.

[6] I. Crinson, "The Governance of Biomedical Science (2): Regulation, Biodata, and Big Data," in The Biomedical Sciences in Society – An Interdisciplinary Analysis. Palgrave Macmillan, Singapore, 2021, pp. 163-181.

[7] C. Chen, H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He, and R. Xia, "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data," Molecular Plant, vol. 13(8), pp. 1194-1202, August 2020.

[8] Y. Li and L. Chen, "Big Biological Data: Challenges and Opportunities," Genomics Proteomics Bioinformatics, vol. 12, pp. 187-189, October 2014.

[9] H. Stevens, Life Out of Sequence: A Data-Driven History of Bioinformatics, University of Chicago Press, 2013.

[10] P. Godfrey-Smith and K. Sterelny, "Biological Information," The Stanford Encyclopedia of Philosophy, E. N. Zalta, Ed., Summer 2016 Edition <https://plato.stanford.edu/archives/sum2016/entries/information-biological/> accessed 05 April 2022.

[11] D. Hallinan and P. De Hert, "Many have it wrong–samples do contain personal data: the data protection regulation as a superior framework to protect donor interests in biobanking and genomic research," in Brent Mittelstadt and Floridi Luciano, Eds. The ethics of biomedical big data, Springer 2016, pp. 119-138.

[12] Z. Yin, H. Lan, G. Tan, M. Lu, A. V. Vasilakos, and W. Liu, "Computing Platforms for Big Biological Data Analytics: Perspectives and Challenges," Computational and Structural Biotechnology Journal, vol. 15, pp. 403-411, 2017.

[13] Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, et al., "Big Data: Astronomical or Genomical," PLoS Biol., vol. 13(7), pp. 1-11, July 2015.

[14] A. Ballantyne, “How should we think about clinical data ownership?” Journal of medical ethics, vol. 46(5), pp. 289-294, 2020.

[15] P. Bronwyn and B. Greenhough, Bioinformation, John Wiley & Sons, 2017.

[16] National Research Council (US) Committee on Frontiers at the Interface of Computing and Biology, “On the Nature of Biological Data,” in J. C. Wooley and H. S. Lin, Eds., Catalyzing Inquiry at the Interface of Computing and Biology, Washington (DC): National Academies Press (US) 2005, <https://www.ncbi.nlm.nih.gov/books/NBK25464/> accessed 05 April 2022.

[17] D. Reinsel, J. Gantz, and J. Rydning, “Data Age 25: The Digitization of the World - From Edge to Core,” IDC White Paper, November 2018 <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> accessed 19 March 2022.

[18] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115-118, January 2017.

[19] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, et al., “Big data: The future of biocuration,” Nature, vol. 455(7209), pp. 47-50, September 2008.

[20] P. Kostkova, H. Brewer, S. de Lusignan, E. Fottrell, B. Goldacre, G. Hart, et al., “Who Owns the Data? Open Data for Healthcare,” Frontiers in Public Health, vol. 4, pp. 1-6, February 2016.

[21] M. Herland, T. M. Khoshgoftaar and R. Wald, “A review of data mining using big data in health informatics,” Journal of Big Data, vol. 1, pp. 1-35, June 2014.

[22] K. Liddell, D. A. Simon, and A. Lucassen, “Patient data ownership: who owns your health?” Journal of Law and Biosciences, vol. 8(2), pp. 1-50, July 2021.

[23] A. de Hingh, “Some reflections on dignity as an alternative legal concept in data protection regulation,” German law journal, vol. 19(5), pp. 1269-1290, October 2018.

[24] United States District Court, Dinerstein v. Google, LLC et al [2020] No: 19 C 4311.

[25] J. Powles and H. Hodson, “Google DeepMind and healthcare in an age of algorithms,” Health Technol., vol. 7, pp. 351-367, March 2017.

[26] BBC Technology, “DeepMind faces legal action over NHS data use,” October 2021, <https://www.bbc.com/news/technology-58761324> accessed 29 March 2022.

[27] M. Wood, “UChicago Medicine collaborates with Google to use machine learning for better health care,” 17 May 2017 <https://www.uchicagomedicine.org/forefront/research-and-discoveries-articles/uchicago-medicine-collaborates-with-google-to-use-machine-learning-for-better-health-care> accessed 01 April 2022.

[28] L. Schencker, “U. of C. Medicine, Google hope to use patterns in patient records to predict health,” Chicago Tribune, 17 May 2017 <https://www.chicagotribune.com/business/ct-google-university-chicago-partnership-0518-biz-20170517-story.html> accessed 15 March 2022.

[29] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, et al., “Scalable and accurate deep learning with electronic health records,” NPJ Digital Medicine, vol. 1, pp. 1-10, May 2018.

[30] T. Minssen, S. Gerke, and C. Shachar, “Is Data Sharing Caring Enough About Patient Privacy? Part I: The Background,” 26 July 2019, <https://blog.petrieflom.law.harvard.edu/2019/07/26/is-data-sharing-caring-enough-about-patient-privacy-part-i-the-background/> accessed 29 March 2022.

[31] I. G. Cohen and M. M. Mello, “Big Data, Big Tech, and Protecting Patient Privacy,” JAMA, vol. 322(12), pp. 1141-1142, September 2019.

[32] L. Schencker, “Judge dismisses lawsuit alleging University of Chicago Medical Center gave Google patient records without consent,” Chicago Tribune, 09 September 2020 <https://www.chicagotribune.com/business/ct-biz-university-of-chicago-google-lawsuit-dismissed-patient-privacy-20200909-lbokttsv6rdc5aeen2q37naady-story.html> accessed 15 March 2022.

[33] J. Becker, “Insufficient Protections for Health Data Privacy: Lessons from Dinerstein v. Google,” 28 September 2020 <https://blog.petrieflom.law.harvard.edu/2020/09/28/dinerstein-google-health-data-privacy/> accessed 19 March 2022.

[34] T. Minssen, S. Gerke, and C. Shachar, “Is Data Sharing Caring Enough About Patient Privacy? Part II: Potential Impact on US Data Sharing Regulations,” 29 July 2019 <https://blog.petrieflom.law.harvard.edu/2019/07/29/privacy-data-sharing-regulation/> accessed 29 March 2022.

[35] D. Wakabayashi, “Google and the University of Chicago Are Sued Over Data Sharing,” The New York Times, 26 June 2019 <https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html> accessed 16 March 2022.

[36] W. N. Price II and I. G. Cohen, “Privacy in the age of medical big data,” Nature, vol. 25, pp. 37-43, January 2019.

[37] S. Gerke, T. Minssen, and G. Cohen, “Ethical and legal challenges of artificial intelligence-driven healthcare,” in Artificial Intelligence in Healthcare, Academic Press, 26 June 2020, pp. 295-236.

[38] Hal Hadson, “Revealed: Google AI has access to huge haul of NHS patient data,” New Scientists, 29 April 2016 <https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/> accessed 29 June 2022.

[39] Information Commissioner’s Office (ICO), RFA0627721 - provision of patient data to DeepMind, 3 July 2017 <https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf> accessed 24 June 2022.

[40] Julia Powles, “Deep Mind’s Latest A.I. Health Breakthrough Has Some Problems,” OneZero, 6 August 2019 <https://onezero.medium.com/deepminds-latest-a-i-health-breakthrough-has-some-problems-5cd14e2c77ef> accessed 24 June 2022.

[41] A. de Hingh and A. R. Lodder, “The role of human dignity in processing (health) data building on the organ trade prohibition,” in EU Internet Law in the Digital Era: Regulation and Enforcement, T. Synodinou, P. Jougleux, C. Markou and T. Prastitou, Eds, Springer, 2020.

[42] A. Bernier, F. Molnár-Gábor, and B. M. Knoppers, “The international data governance landscape,” Journal of Law and the Biosciences, vol. 9(1), pp. 1-45, January-June 2022.

[43] A. Mcmahon, A. Buyx, and B. Prainsack, “Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in The Governance of Data Use in Medicine and Beyond,” Medical Law Review, vol. 28(1), pp. 155-182, August 2019.

[44] B. Prainsack and A. Buyx, Solidarity in Biomedicine, Cambridge University Press, 2017.

[45] Regulation (EU) 2018/1724 of the The European Parliament and of the Council on European data governance and amending (Data Governance Act), 4 May 2022, 2020/0340 (COD) <https://data.consilium.europa.eu/doc/document/PE-85-2021-INIT/en/pdf> accessed 05 July 2022.

[46] Masha Shabani, “The Data Governance Act and the EU’s move towards facilitating data sharing,” Mol Syst Biol., Commentary, 17: e10229, pp. 1-3, 2021.

[47] L. Floridi, “Open Data, Data Protection, and Group Privacy,” Philos Technol., vol. 27, pp. 1-3, February 2014.

[48] L. Taylor, L. Floridi, and B. van der Sloot, Group Privacy: New Challenges of Data Technologies, Dordrecht: Springer, 2017.

[49] S. Viljoen, “Democratic Data: A Relational Theory for Data Governance,” Yale Law Journal, vol. 131, pp. 1-67, November 2020.

[50] A. E. Waldman, “Privacy Law’s False Promise,” Washington University Law Review, vol. 97(3), pp. 773-834, 2020.

[51] The White House, Draft memorandum for the Heads of Executive Departments and Agencies, “Guidance for regulation of artificial intelligence applications,” January 2020 <https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf> accessed 26 June 2022.

[52] The White House, “OSTP’s Continuing Work on AI Technology and Uses that Can Benefit Us All,” OSTP Blog 03 February 2022 <https://www.whitehouse.gov/ostp/news-updates/2022/02/03/ostps-continuing-work-on-ai-technology-and-uses-that-can-benefit-us-all/> accessed 04 July 2022.

[53] Independent High-Level Expert Group on Artificial Intelligence Set Up by the European Commission, “The Assessment List for Trustworthy Artificial Intelligence (ALTAI) - for self assessment,” 17 July 2020 <https://digital-strategy.ec.europa.eu/en/library/assessment-list-

trustworthy-artificial-intelligence-altai-self-assessment> accessed 27 June 2022.

[54] Independent High-Level Expert Group on Artificial Intelligence Set Up by the European Commission, "Ethics Guidelines for Trustworthy AI," 8 April 2019 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 27 June 2022.

[55] Data Protection Impact Assessment (DPIA), "How to conduct a Data Protection Impact Assessment," <https://gdpr.eu/data-protection-impact-assessment-template/> accessed 30 June 2022.