



DATAETHICS Intellectual Output 1

Recommendations and analysis of the competences and evaluation of expertise



Co-funded by the
Erasmus+ Programme
of the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

INTELLECTUAL OUTPUT 1

Diagnosis and Recommendations

Authors:

József Kovács MD, PhD, DSc, Professor of Bioethics, Head of the Department of Bioethics at the Institute of Behavioural Sciences, Semmelweis University, Budapest, Hungary

Orsolya Márta Péter JD, PhD, LLM, Senior Lecturer in Bioethics at the Institute of Behavioural Sciences, Semmelweis University, Budapest, Hungary

Gábor Attila Tóth JD, PhD, Associate Professor of Bioethics at the Institute of Behavioural Sciences, Semmelweis University, Budapest, Hungary

Csaba Bödör MD, PhD, DSc, Head of Molecular Diagnostics, Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

Bence Bártai MD, PhD student, Semmelweis University, Budapest, Hungary

- with DATAETHICS Consortium

EUROLIFE DATAETHICS

<https://www.dataethics-eurolife.eu/>

Index

<i>I. Background and Rationale of the DATAETHICS Project</i>	3
<i>II. Aims and Goals of the IO1 Task Force</i>	4
<i>III. Introduction: An Overview of the Ethical Risks and Benefits Associated with Big Data in the Field of Medicine (J. Kovács)</i>	4
III.1. Discrimination and Bias.....	5
III.2. Privacy Issues	5
III.3. Anonymity and a Changing Concept of Informed Consent	5
III.4. Scientific Research	6
III.5. Conclusions	7
<i>IV. Analyses of the EUROLIFE DATAETHICS Activities</i>	9
IV.1. Ethics Education for Medical Students in the European Union	9
IV.2. Analysis of the Student Surveys from 2020 (128 Students) (Cs. Bödör, B. Bártai)	10
IV.3. Analysis of Documents from the Virtual Exercise (G. A. Tóth)	18
IV.3.1. Conceiving a Case Story Addressing Ethical Issues.....	18
IV.3.2. Definitions, Interpretations and Core Ethical Issues Identified by the Students	20
IV.3.3. Understanding and Recognising Ethical Issues in the Context of Biomedical Big Data in the Student Presentations.....	27
IV.4. Analysis of Student Competences Based on the Case Story Presentation in the Online Pilot School (O.M. Péter).....	30
IV.4.1. The Assignment.....	30
IV.4.2. Execution of the Assignment.....	31
IV.4.3. Student Feedback on Skills and Competences.....	37
<i>V. Conclusions and Recommendations</i>	41

I. Background and Rationale of the DATAETHICS Project

DATAETHICS - Changing Landscapes in the Health and Life Sciences: Ethical Challenges of Big Data, is a transnational project aiming at updating the conceptual and ethical considerations that arise from Biomedical Big Data (BBD) collection and use.

From the completion of the Human Genome Project onwards, the advent of technological advances in the collection and analysis of BBD has considerably outpaced any conceptual and moral considerations that inevitably follow. Thus, the generation of such large datasets is no longer the major challenge, but rather the management, handling and interpretation thereof. The coupling of ethical considerations with scientific ones in learning structures is now a necessary and unavoidable step in order to prevent us from adopting a solely technocratic approach toward BBD.

This project supports EU and national policies in the innovation and use of BBD reflected in initiatives for digital transformation, as illustrated by the recent proposal for the Digital Europe Programme (2021-2027) and aligned to the European Economic and Social Committee's report on the ethics of BBD.

Given that the pace of data collection and generation in biomedical sciences (e.g. the advent of genomics technologies; high throughput imaging; patient cohort data collections etc.) has far exceeded the pace by which students are currently being educated on these topics, we have identified a substantial need for updated study materials and novel educational approaches in order to bridge this gap. With the novel practical tools, contents and guidelines offered herein, the DATAETHICS project aims to generate improved and durable availability of training programmes with real-life examples, which will allow us to offer broad and profound skills and knowledge that both students and instructors currently might need in the field of life sciences. Furthermore, DATAETHICS will facilitate the unification of learning criteria and will open a window of opportunity for continuous improvement in the delivery of quality teaching in Higher Education.

The project is implemented by a consortium of nine top-tier European academic institutions and medical centres from nine countries, each with a worldwide reputation in research excellence. Complemented by the engagement of a large network of associate partners, who will participate in the process of knowledge production and dissemination, the project will additionally operate in order to further enhance industry-university cooperation.

The improvement of medical knowledge and clinical care lies at the centre of EU policy. To this end, DATAETHICS will expedite cooperation between EU countries in order to identify best practices for BBD.

II. Aims and Goals of the IO1 Task Force

As a first step, the task force intended to analyse the currently existing knowledge of ethical considerations in connection with Biomedical Big Data (BBD) across the partner institutions, primarily as reflected in the feedback provided by the prospective and, later, effective participants of the first initiatives within the framework of the project. Such analyses would enable us to develop an attractive education and training programme, in line with the identified needs and expectations of both students and university staff at the partner institutions.

Hence, the IO1 objectives include certain analyses of the initial EUROLIFE DATAETHICS activities that are to serve as an input for the future education and training programmes to be developed to such goal by the consortium. These activities are as follows:

- i) Analysis of the data provided by the first overall student survey that served as a starting point for the development of the pilot training programme.
- ii) Analysis of the data provided in connection with the first Virtual Exercise, developed on the basis of the initial feedback provided by prospective graduate and postgraduate student participants.
- iii) Analysis of the achievements of the first Online Pilot School.

III. Introduction: An Overview of the Ethical Risks and Benefits Associated with Big Data in the Field of Medicine (J. Kovács)

The term “Big Data” describes a paradigm shift from “Small Data” (that we traditionally had before the introduction of contemporary digital technology) to an utterly new concept. “Big Data” is usually defined by the “three V’s”: “Volume” (very large amount of data); “Variety” (the data to be analysed are heterogeneous, representing different types of data); and “Velocity” (the data can be accessed and analysed with extreme speed)¹. “Biomedical Big Data” (BBD) or, alternatively, “Health Big Data” is a somewhat narrower term, including all data that refer to our health status and/or our disease risks.

In this introduction we will start from the premise that personal health data are highly sensitive ones and deserve a high level of protection². On the other hand, data are particularly valuable when they are collected from a vast amount of people, to be combined and analysed at a later date³.

Although medical confidentiality had already been an important ethical injunction in the Hippocratic Oath, medicine always tried to collect and store patient data, as it was held that the more a physician knows about their patient, the better. In this era of pre-electronic health records the awareness of the risks that the storage of such data represent was minimal⁴, and the main effort was aimed at preventing unwanted data disclosure. Accordingly, our ethical sensibilities were developed to protect our privacy in the “Small Data” world. The fast development of modern technologies in the 21st century, however, resulted in the equally swift transition into the era of “Big Data”.

III.1. Discrimination and Bias

Big Data may bring improvement in health care by offering personalized medicine, precision medicine, higher quality of care and a litany of other advantages⁵. On the other hand, “Big Data is big business”. Data brokers, i.e. companies that collect, analyse and sell personal information of consumers, often classify consumers into sensitive health-related categories based upon their interests, e.g. “Expectant parent”, “Diabetes interest”, etc.⁶ Thereby Big Data can also be used to identify who these individuals may be, and currently there is little protection to prevent these people from being discriminated against by interested parties such as employers or insurance companies. Strong anti-discrimination laws should protect individuals who are healthy at present, but are predicted to become disabled in the future⁷.

Another important problem is that Big Data does not include everybody. The richer somebody is, the more (s)he has access to the technologies that constitute the sources of Big Data. In this way Big Data does not represent the natural diversity of a given population. The data of ethnic minorities, data of people of low socio-economic status, who are not present on social media and/or do not use credit cards or smart phones; in short, who are poor and disadvantaged are usually missing from the sources of Big Data. This means that these people are left out of (the benefits offered by) Big Data, thus creating biased datasets – which is particularly damaging, because the less privileged are customarily the ones most in need for medical care⁸.

III.2. Privacy Issues

In many countries the current legal framework for healthcare data protection applies only to data created by health care. However, today, only a fraction of BBD is generated by health care professionals. We use smart phones, smart watches, fitness trackers, and other health monitoring devices that can record almost every aspect of our bodily functioning and daily behaviour. These data are customarily analysed by AI, i.e. machine learning and predictive analytics based on data mining⁹.

On the one hand, these technologies empower the patients who can monitor their own health without the help of their doctors¹⁰. On the other hand, these devices render the patient more vulnerable, because (s)he becomes transparent to any party that has access to these data. Mobile phones and health applications are notoriously insecure from the point of view of privacy¹¹.

However, the EU General Data Protection Regulation (GDPR), and specifically Article 4(15), is an exception from the rule that legal regulations refer only to health data created by health care institutions and health care professionals. The GDPR has a specific category of “*data concerning health*”, which includes “*personal data related to the physical or mental health of a natural person which reveal information about his or her health status.*” Disease risk is included in health data as well¹². Thus, the health data collected by non-medical devices such as smart watches and other wearables are provided with a higher level of protection by the GDPR.

III.3. Anonymity and a Changing Concept of Informed Consent

In the environment of Big Data, apparently, the informed consent rule of Small Data bioethics can no longer protect our sensitive health data¹³. Traditionally, in the era of Small Data, de-identifying (anonymising) personal data was the tool to protect privacy and prevent unwanted data disclosure. The traditional approach to de-identifying data was to remove all off-line identifiers, like names or addresses. However, in the era of modern analytics, this is not enough, because even anonymous people can be classified and targeted according to their online behaviours and preferences¹⁴.

Furthermore, in the era of Big Data, the vast amount of available data makes the re-identification of individuals still possible. Today, nobody can “hide in the crowd.”¹⁵ The de-identified data, extracted from Electronic Health Records (EHRs) of patients, patient registries and databases used in health care can often be re-identified and combined with the information originating from social media and health apps¹⁶.

Up to now, legal regulations made a sharp distinction between identified data and non-identified or de-identified data. They required informed consent of the data subject to use her/his data for research, even if obtaining informed consent years after the data had been created proved to be difficult or almost impossible. By contrast, doing research on de-identified or non-identified data was possible without the informed consent of the data subject. This distinction, however, seems to be obsolete today, where almost every de-identified or non-identified data can theoretically be re-identified. On the other hand, requiring informed consent for research on identified data, which have been created years earlier, is very difficult. In our era of Big Data, millions of research subjects should be contacted, which would make effective research impossible. Some therefore argue that the data contained in EHRs should be allowed to be used for research purposes without obtaining informed consent¹⁷.

Such moral (and legal) duty to share healthcare data wouldn't mean that patients shouldn't be informed about the utilisation of their data. They should be aware of the use of EHR data, but such usage shouldn't require their informed consent¹⁸. The most important precondition should be the protection of the privacy of the patient. The data in the EHRs should be de-identified, and reasonable precaution should be made in order to prevent re-identification.

III.4. Scientific Research

The Declaration of Helsinki defines medical research as “*involving human subjects, including research on identifiable human material and data.*”¹⁹ The somewhat stricter US federal Common Rule considers research on patients' data to be human research only if the data are private and individually identifiable²⁰.

Using these definitions, Big Data health research usually does not equal human research, provided it does not involve intervention or interaction with patients specifically with the purpose to obtain their data because the information obtained by it is not private and not individually identifiable²¹. Still, Big Data health research projects are routinely submitted to Research Ethics Committees (RECs) in Europe, and Institutional Review Boards (IRBs) in the USA, because currently there is no alternative system to review and monitor Big Data health research projects from the viewpoint of ethics²².

These committees — in accordance with the generally accepted standards of Western research ethics — may require the informed consent of the research subjects, which is practically impossible in the case of Big Data, where the data of millions of patients are to be analyzed. The rules of informed consent were originally created for research carried out on small datasets, where protecting the privacy of patients required informed consent. Big Data health research requires a different standard.

In the era of Big Data almost every anonymised or de-identified data are potentially re-identifiable. It is usually left to the research ethics committees to decide whether this fact renders Big Data health research human research or less. If Big Data health research is to be considered human research, then the requirement of individual informed consent will make the research practically impossible to carry out. On the other hand, analysing publicly available information, even if it is individually identifiable, is not human research if the information is not private²³.

The risk of research with data is informational risk. The disclosure of individual and private information may cause harm to research subjects²⁴. According to the common rule in the US, if a research presents only minimal risk to the subjects, then the IRBs can permit a waiver of informed consent if it is considered necessary. Because Big Data health research cannot be carried out without the waiver of informed consent, such research can be permitted only if it involves minimal informational risks. One way of defining minimal risk is using the „daily life risk standard“ according to which a risk is minimal, if it is not greater than the level of risk encountered in daily life²⁵. However, the problem with this approach is that health data are highly sensitive data, the disclosure whereof can be seriously harmful to subjects; consequently, they cannot be regarded as involving only minimal risk. If the risk, however, is more than minimal, the waiver of informed consent cannot be permitted, which makes Big Data health research impossible to carry out.

III.5. Conclusions

As it is illustrated by the foregoing, modern technologies have ushered in a new era of medical possibilities, but they also pose a host of complex challenges for medical ethics. These challenges arise from the rapid evolution of technologies like, among others, data analytics.

With the proliferation of electronic health records, wearable devices, and telemedicine platforms, maintaining patient privacy and securing sensitive medical data has become a significant concern. Ensuring the privacy and security of this data against breaches, unauthorized access, and misuse is a significant ethical concern.

Patient-generated health data collected through wearables and mobile applications can provide valuable insights. However, the ownership, control, and usage of this data by healthcare providers, technology companies, and researchers require clear ethical guidelines.

Also, while modern technologies have the potential to revolutionize healthcare, there is a risk that they may only be accessible to those who can afford them, exacerbating existing healthcare disparities. Access to advanced medical technologies and treatments, developed – among others – by

the use of Biomedical Big Data might not be evenly distributed, leading to concerns about unequal access to benefits.

Addressing these challenges requires collaboration among ethicists, healthcare professionals, technologists, policymakers, and the public. Striking a balance between technological innovation and ethical considerations is essential to ensure that modern medical advancements are deployed in ways that prioritize patient well-being, justice, and equity. Developing guidelines, regulations, and ethical frameworks that account for the complexities of modern medical technologies is essential to ensure that patient care remains at the forefront of technological progress. As part of such process, the continued education of medical students in ethics in general, and in the modern challenges outlined hereinabove, is therefore inevitable.

IV. Analyses of the EUROLIFE DATAETHICS Activities

IV.1. Ethics Education for Medical Students in the European Union

Ethics education for medical students is generally considered an important aspect of medical training in the European Union, but the specifics vary from one country to another. The EU does not have a unified policy regarding ethics education for medical students, as education policies and regulations are determined at the national level by each Member State. Many EU countries include medical ethics as a mandatory component of the medical curriculum, recognizing the importance of preparing future healthcare professionals to navigate complex ethical dilemmas. However, the extent and depth of ethics education can differ from institution to institution.

While approaches may vary among EU Member States and medical schools, apparently there are certain common themes in how medical ethics is customarily taught. Bioethics/medical ethics is typically integrated into the medical curriculum throughout various stages of education. Ethics education typically begins with a theoretical foundation that covers ethical theories, principles (autonomy, beneficence, non-maleficence, justice), and fundamental ethical concepts. Ethics is often taught in conjunction with relevant clinical cases and scenarios to provide practical context. As an acknowledged teaching method in certain medical schools, students may engage in case-based learning, analyzing real or fictional scenarios to identify ethical issues, consider different perspectives, and propose ethical solutions.

Furthermore, ethics education often involves collaboration with experts from various disciplines, such as philosophy, law, and social sciences. This interdisciplinary approach enriches discussions on ethical principles and their application to medical practice.

In several medical schools emphasis is placed on teaching effective communication skills and the importance of obtaining informed consent from patients, and education emphasizes patient-centered care, focusing on effective communication, respect for patients' autonomy, and shared decision-making. This aligns with the EU's commitment to patient rights and dignity. Also, medical students are often encouraged to engage in self-reflection and self-awareness activities to better understand their own values, biases, and potential conflicts of interest that could impact their ethical decision-making.

Ethics education usually covers ethical considerations in medical research as well, including informed consent, privacy, data protection, and the use of human subjects. This is aimed to ensure that students understand the ethical dimensions of scientific inquiry.

However, the recently developed modern technologies – including the creation of Biomedical Big Data – have introduced a range of ethical challenges for medical students, but also for practicing medical professionals, and bioethicists. As outlined hereinabove, the rapid growth in the volume, complexity, and accessibility of healthcare data poses unique obstacles that the medical profession – including medical students – must address, in cooperation with ethicists, sociologists, lawyers and philosophers.

IV.2. Analysis of the Student Surveys from 2020 (128 Students) (Cs. Bödör, B. Bártai)

With regard to the foregoing, as a first step towards developing a specific programme, it was considered necessary to obtain insights from the students of the EUROLIFE partner universities on their understanding of the concept and ethical issues connected to Biomedical Big Data. To such goal an online questionnaire was distributed among the students with the following questions:

General questions:

1.	Do you understand what Big Data in Life/Biomedical Sciences/Medicine is? (Yes/No)
2.	(If answered "Yes" in Question 1): Please give one example in one sentence. (<i>maximum one sentence text</i>)
3.	In your opinion, what are the BENEFITS and RISKS the Big Data/technology entails for life/biomedical science professionals, patients, and society? (<i>column for Benefits and column for Risks: List maximum 5 benefits and/or risks.</i>)
4.	Do you think that new ethical considerations arise when analyzing or integrating different data from individuals in order to gain novel insights? (Yes/No)
5.	Do you think that your institute's curriculum sufficiently covers ethics issues in the context of data collection/use/interpretation and application? (Yes/No)
6.	(If answered "Yes" in Q5): Please provide a maximum two-sentence description of the course/activity. (maximum two-sentence text)

Collaboration of clinical and data scientists - implications of data evaluation and results

1.	Do you see the importance of a skillset to evaluate the methods of interpretation of patient's data and/or the resulting findings in data from that patient? (Yes/No)
2.	Do you think patients should be informed on unexpected implications of the datasets collected from them by a data scientist? (Yes/No)
3.	(If answered "Yes" in Q2): Who should communicate these findings to the patient? Please name the professional. (Maximum one sentence text) Would <u>you</u> currently feel adequately prepared to do so? (Yes/No)

Uniformity in the consideration of Big Data

1.	Do you know of a common environment/platform in Life Sciences/Medicine for data scientists (students, researchers and clinicians) providing guidelines and tools for a safe and compliant data collection/use/interpretation? (Yes/No)
2.	(If answered "Yes" in Q1): Please name this environment/platform. (Maximum one sentence text)
3.	(If answered "No" in Q1): Do you think that there should be a uniform ethical code when considering Big Data in an academic or medical versus a non-academic or medical/commercial context? (Yes/No)

Practical:

1.	Do you think that you could advance your understanding of Big Data in Life/Biomedical Sciences/Medicine via structured interdisciplinary and international/intercultural cooperation? (Yes/No)
2.	Would you consider taking part in extracurricular activities/courses on the topic of Big Data and Ethics? (Yes/No)
3.	Please explain in one-sentence what sort of format would you imagine such a course/activity

	to have? (maximum one-sentence text)
4.	4. How would you value the importance of personal meetings with fellow students and teachers in the context of such a course/activity? (graph for answer - scale from 1 to 5)
5.	What duration would you consider appropriate for such a learning course/activity? (list to select from 1 day, 2 days, 3 days, 4 days, 5 days, 1 week, 2 weeks, 3 weeks, 1 month, 2 months)
6.	What qualification/recognition would you expect following completion of such a course/activity? (maximum 1 sentence text)
7.	What would be the expected impact of such an activity/course on your personal/professional life? (maximum 1 sentence text)

Analysis

In total, 128 surveys were filled and returned by the students expressing interest in the EUROLIFE DATAETHICS project (Figure 1).

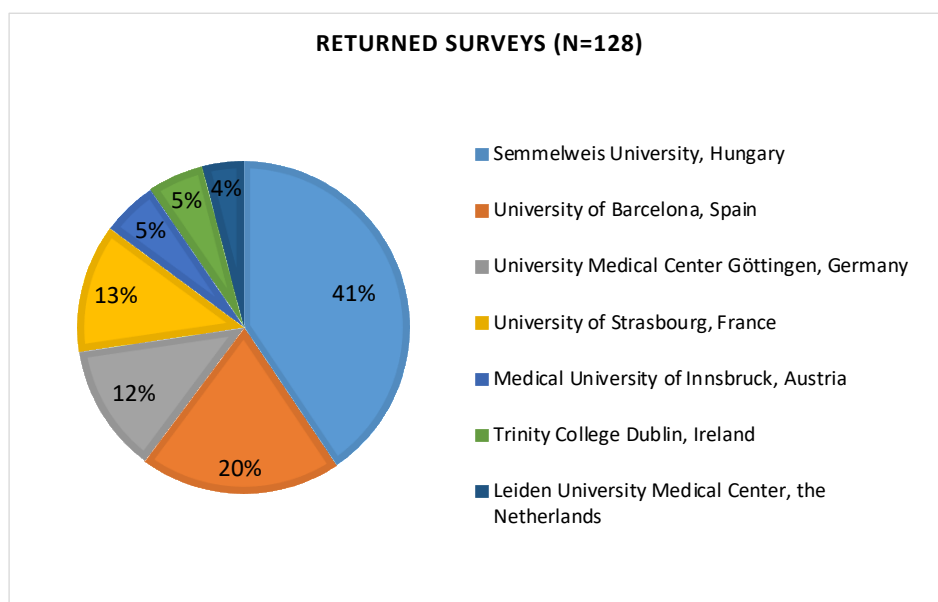


Figure 1: Distribution of the returned surveys based on the participating partner universities.

Questions 1 and 2

1) Do you understand what Big Data in Life/Biomedical Sciences/Medicine is? (Yes/No)

2) (If answered “Yes” in Question 1): Please give one example in one sentence. (Maximum one sentence text)

54% of the students had correct understanding of Biomedical Big Data, 38% reported that they did not understand what Big Data in biomedical sciences represents, with 8% of the students claiming understanding, however failing to provide a relevant example for Biomedical Big Data (Figure 2).

DO YOU UNDERSTAND WHAT BIG DATA IN LIFE/BIO MEDICAL SCIENCES IS?

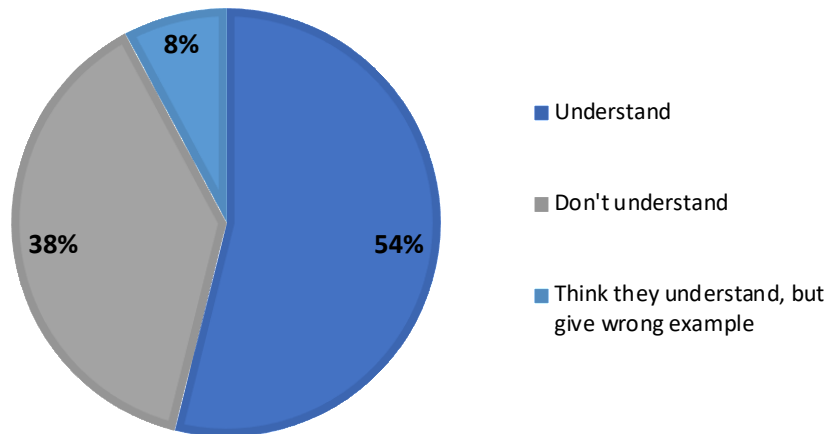


Figure 2: Distribution of answers on understanding the concept of Biomedical Big Data.

Question 3

In your opinion, what are the BENEFITS and RISKS the Big Data/technology entails for life/biomedical science professionals, patients, and society? (*column for Benefits and column for Risks: List **maximum 5 benefits and/or risks.***)

The benefits and risks associated with Biomedical Big Data reported by the students are summarized in word clouds (Figures 3 and 4)

15/11/2021, 09:40

Word Art

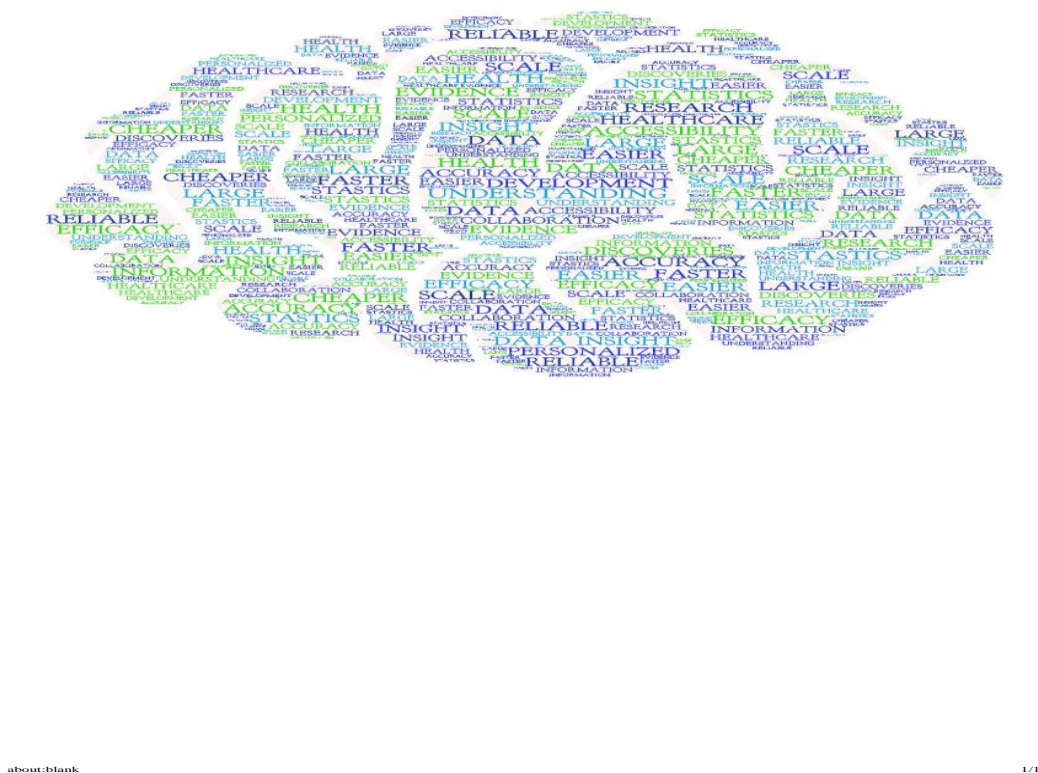


Figure 3. The benefits of Big Data in biomedical sciences based on the student surveys. Answers were grouped into categories and word numbers are proportionate to their occurrence.

Question 5

Do you think that your institute's curriculum sufficiently covers ethics issues in the context of data collection/use/interpretation and application? (Yes/No)

80% of students felt that their institutions' curricula do not sufficiently cover ethical issues connected with Biomedical Big Data (Figure 6).

**DO YOU THINK THAT YOUR INSTITUTE'S CURRICULUM SUFFICIENTLY
COVERS ETHICS ISSUES IN THE CONTEXT OF DATA
COLLECTION/USE/INTERPRETATION AND APPLICATION?**

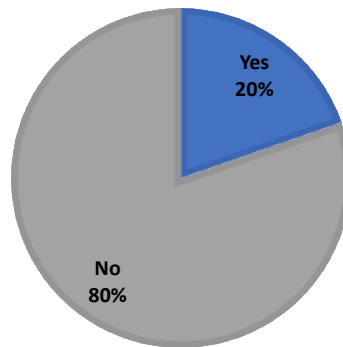


Figure 6. Distribution of answers for Question 5.

Collaboration

Questions 2 and 3

2) Do you think patients should be informed on unexpected implications of the datasets collected from them by a data scientist? (Yes/No)

3) (If answered "Yes" in Q2): Who should communicate these findings to the patient? Please name the professional. (Maximum one sentence text) Would you currently feel adequately prepared to do so? (Yes/No)

79% of the students felt that patients should be informed on unexpected findings arising from analysis of their data, however 73% of them do not feel adequately prepared to do so (Figures 7 and 8).

DO YOU THINK PATIENTS SHOULD BE INFORMED ON UNEXPECTED IMPLICATIONS OF THE DATASETS COLLECTED FROM THEM BY A DATA SCIENTIST?

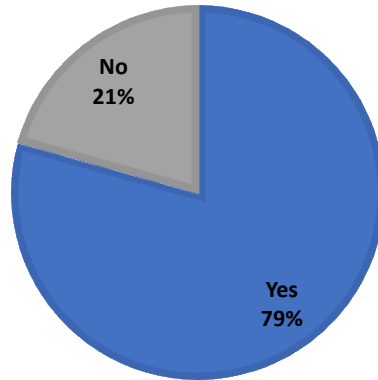


Figure 7. Distribution of answers for Question 2.

WOULD YOU CURRENTLY FEEL ADEQUATELY PREPARED TO DO SO?

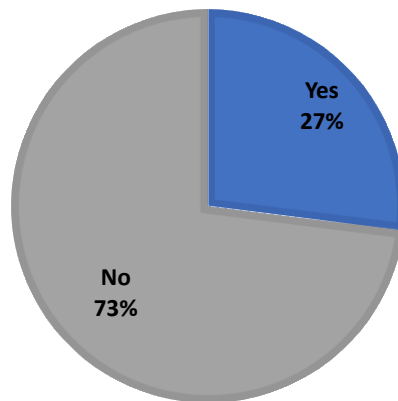


Figure 8. Distribution of answers for Question 3.

Practical:

Questions 1-3

- 1) Do you think that you could advance your understanding of Big Data in Life/Biomedical Sciences/Medicine via structured interdisciplinary and international/intercultural cooperation?
- 2) Would you consider taking part in extracurricular activities/courses on the topic of Big Data and Ethics?
- 3) Please explain in one-sentence what sort of format would you imagine such a course/activity to have?

98% of the students reported that their understanding of Biomedical Big Data would advance from a structured international cooperative education program (Figure 9). 85% of the students expressed interest in participating in this type of education programme (Figure 10).

The preferred formats of these education activities are summarized in Figure 11.

DO YOU THINK THAT YOU COULD ADVANCE YOUR UNDERSTANDING OF BIG DATA IN LIFE/BIOMEDICAL SCIENCES/MEDICINE VIA STRUCTURED INTERDISCIPLINARY AND INTERNATIONAL/INTERCULTURAL COOPERATION?

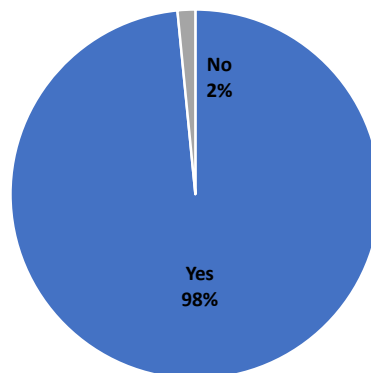


Figure 9. Distribution of answers for Question 1.

WOULD YOU CONSIDER TAKING PART IN EXTRACURRICULAR ACTIVITIES/COURSES ON THE TOPIC OF BIG DATA AND ETHICS?

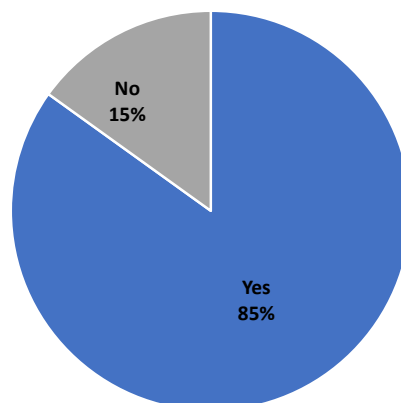


Figure 10. Distribution of answers for Question 2.

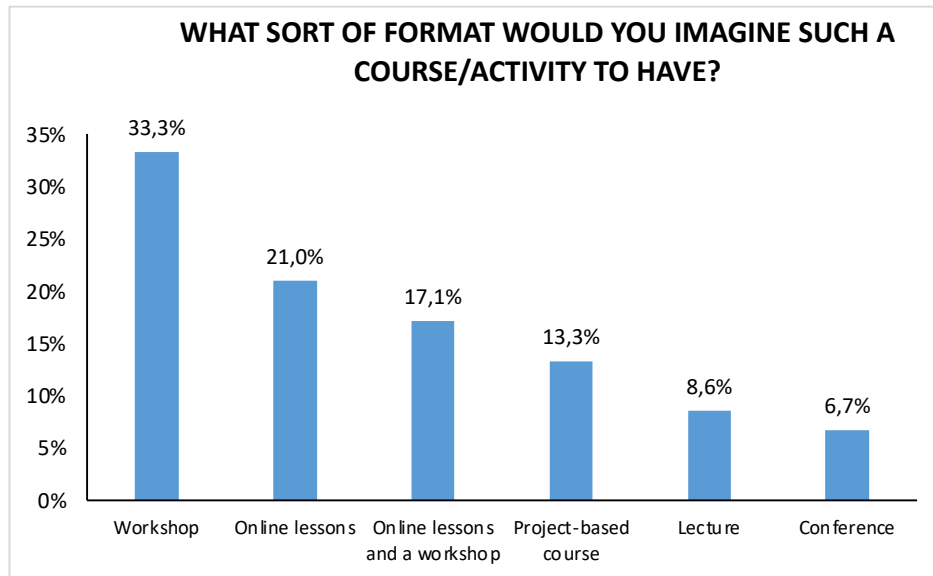


Figure 11. Distribution of answers for Question 3.

Conclusions of the student survey:

- 46% of the students did not have correct understanding of the Biomedical Big Data concept.
- 91% of the students were aware of the new ethical considerations arising in connection with use of Biomedical Big Data.
- 80% of students felt that their institutions' curricula do not sufficiently cover ethical issues connected with Biomedical Big Data.
- 98% of the students reported that their understanding of Biomedical Big Data would advance from a structured international cooperative education program.
- 85% of the students expressed interest in participating in this type of education program preferably in workshop or online lesson format.

Such findings are in line with the premises outlined in Section IV.1. hereinabove (*"Ethics Education for Medical Students in the European Union"*), i.e. medical students in the EU encounter some level of ethics education during their training; however, the specific content, format, and extent of ethics education varies among institutions. In addition to such differences, as far as modern technological developments and, more specifically, Biomedical Big Data are concerned, the answers reflected that the respective concerns and challenges outlined hereinabove are valid and real.

Quite understandably, Biomedical Big Data obviously presents a significant and novel challenge for medical students due to its sheer volume, complexity, and potential impact on healthcare. As the field of medicine becomes increasingly data-driven, medical students – in a manner not entirely different from what practising medical professionals and researches experience – obviously feel that their knowledge and level of preparedness is currently insufficient for properly handling and utilizing large-scale health data. At the same time they – also quite obviously – feel the need to develop skills in data management to effectively navigate and utilize these diverse datasets as well, including the ethical considerations and privacy regulations associated with handling and accessing patient data to ensure data security and patient confidentiality.

IV.3. Analysis of Documents from the Virtual Exercise (G. A. Tóth)

The virtual exercise was conceived by the consortium in cooperation with the Association for Values and Knowledge Education (AVaKE). It was agreed that Values and Knowledge Education (VaKE) would be the method of choice to be referred to in the DATAETHICS project, in combination with other methods. Neither the experts nor the students have had any previous experience with VaKE. While most ethics experts are familiar with the task of conceiving and working with dilemmas, this method is less commonly used by biomedical scientists or practicing clinicians, therefore students – depending on how their home institution approaches the teaching of ethics and what methods are preferred – may be less experienced in learning from dilemmas. Hence, the goal was to obtain experience about VaKE, as well as working with dilemmas on the level of the students, ultimately in order to continue adapting and utilising the VaKE principles within the DATAETHICS project.

As far as the VaKE teaching method is concerned, the essence thereof – in the words of its creators J.-L. Patry and S. Weyringer²⁶ – is as follows: *„In a VaKE process the learners are confronted with a dilemma addressing moral issues which they discuss; if the dilemma is conceived appropriately, it will trigger knowledge questions, which are then answered by the learners through an information search in relevant sources available. In a VaKE process, then, the learners perform a double research process: ethical justification (research with respect to argumentation in favor or against moral values in the given dilemma situation), and empirical (research with respect to looking at empirical evidence necessary for their argumentation).”* Since it was originally developed in the early 2000s, it has been implemented in various educational contexts, such as the training of teachers, secondary school students, and healthcare professionals.

The virtual exercise consisted of four tasks, to be accomplished by the students.

IV.3.1. Conceiving a Case Story Addressing Ethical Issues

The first task was to conceive a case story addressing ethical issues, with the assistance of a questionnaire. This task had the goal, for the researchers, to identify certain core topics of BBD with ethical relevance, from the point of view of the participating students, who had no previous experience with moral dilemmas. The task provides a first insight into the sensitivity of the students for BBD issues and for dilemmas that can be involved. One of the proposed case stories was to be chosen and adapted for discussion in the next tasks, which consist in approximations of VaKE discussions and presentation.

The student questionnaire was provided with the following structure.

- Sections 1-12: Understanding and recognition of ethical issues in the context of Biomedical Big data based on an individual case story;
- Section 13: Definition of Biomedical Big Data;
- Sections 14-17: Understanding of principles and rules in ethics in general and in the context of Biomedical Big Data.

The instructions were as follows.

Have you experienced or can imagine ethical issues with Big Data in the field of biomedicine (Biomedical Big Data – BBD)? Have you ever made an experience with the use and misuse of data, and that a decision had to be made in this regard which was somehow problematic, ambivalent, not trivial or challenging? Please describe such a situation. If you have not experienced such a situation, please imagine one.

Four students per university (from nine EUROLIFE/DATAETHICS universities; one university had only three participants) were planned to participate. The participants were recruited by the DATAETHICS representative of the respective university. It is unknown what further information and instruction they received from the local experts. Although collaborative work was not prohibited, the task was conceived as an individual one, with an individual product to be handed in. Finally, 25 students submitted their answers.

The questionnaire was structured as follows:

(1) Short description of the case story

(2) Describe the protagonist who has to take the decision, as far as it is important for the story.

(3) Describe the central issues and parameters of the situation

(4) Who else is involved?

(5) What roles do these persons have? In what regard are they involved? (The persons and their respective roles were to be itemised in a table.)

(6) What has happened until this point?

(7) What other conditions are important?

(8) Describe the decision the protagonist has to take, and what options the protagonist has: What can the protagonist do? (The various hypothetical options were to be itemised in a table – Option 1, Option 2, further options.)

(9) Describe how the people are affected by the decision. (The possible consequences were to be itemised a table in regards of each individual.)

(10) In the case that you experienced yourself this situation, how did the protagonist decide, and why? In the case that you imagined the story, how should the protagonist decide, in your view, and why? (The possible decisions and the respective reasons were to be itemised in a table.)

(11) In the case that you experienced this situation: What were the consequences of the decision? In the case that you imagined the story and the decision: What do you think are the probable consequences?

(12) In what regard is BBD relevant or addressed in this story?

(13) How do you define BBD in general?

(14) How do you define Ethics in general?

(15) In your opinion, what principles or rules are most important in Ethics? (Up to five, in hierarchical order according to their importance in general, not restricted to biomedicine or Big Data.) Why do you think these principles are important? (The order of priorities, together with the respective explanations, was to be itemised in a table.)

(16) In your opinion, what principles or rules are most important that are particularly relevant in biomedicine? (Up to five, in hierarchical order according to their importance in biomedicine.) Why do you think these principles are important? You can take up the same principles as above (in the same or a different order), or add new ones. (The order of priorities, together with the respective explanations, was to be itemised in a table.)

(17) In your opinion, what principles or rules are most important that are particularly relevant for big data? (Up to five, in hierarchical order according to their importance with respect to big data.) Why do you think these principles are important? You can take up the same principles as above (in the same or a different order), or add new ones. (The order of priorities, together with the respective explanations, was to be itemised in a table.)

For the purpose of the following summary, PDF documents from individual students served as primary sources, and the analysis of the virtual exercise by the AVaKE experts served as additional source.

IV.3.2. Definitions, Interpretations and Core Ethical Issues Identified by the Students

A. Definitions of Biomedical Big Data by the students (Section 13)

As a general concept accepted by contemporary science, Biomedical Big Data refers to the analysis of aggregated datasets in order to advance medical knowledge and improve the quality of clinical care. Big Data is typically characterized by three defining features, known as the “3Vs”: Volume, Variety and Velocity.

From among the participants, 23 students submitted their respective definitions of BBD. All definitions included certain essential elements of the aforementioned definition of BBD. Many definitions had two or three relevant keywords, while others offered detailed descriptions with examples.

Most of the students were aware that BBD covers biological and clinical data. They understood that the subject matter of BBD is biomedical and healthcare informatics research.

Almost all of the definitions referred to the volume of data in some manner (*large, great, or huge amounts, big or large scale* etc.). The variety of data types and structures is mentioned only in a couple of definitions, either as a general term (e.g. *variety or complexity*) or as an illustration (e.g. *brain imaging, genome, proteome or genetic data and digital health records*). Nevertheless, the third characteristic of Big Data, i.e. velocity - referring to the fast speed of producing and processing data - hardly emerged from the definitions.

About half of the definitions involved an explanation of BBD in terms of the purpose it serves. Among the answers we find that the datasets aim to “*create new methodologies to advance in biomedical sciences*”; “*identify potential diseases and linked to variants or mutations.*”

Beyond the defining characteristics some answers included normative elements, e.g. that BBD “*imposes a threat to privacy,*” while others extended the scope of BBD to “*pseudo-alive subjects (not necessarily restricted to our species).*” Even though it seems to be generally known that BBD is too complex to be dealt with traditional data-processing methods, among the rare misconceptions that featured in the answers provided by the students we could find that BBD is “*generated by health care professionals.*”

In sum, the definitions show that the core meaning and significance of BBD were known by the majority of the students who contributed to the questionnaire. This finding also reflects that the participants were already well aware of the complex phenomenon which is BBD, and they made individual efforts in order to familiarise themselves with the associated fundamentals.

B. Understanding of principles and rules in ethics in general and in the context of Biomedical Big Data (Sections 14-17)

In this section 24 students offered different definitions of ethics. The majority of definitions provided two or three relevant keywords, while others gave detailed descriptions with examples. The answers reflect that the word “ethics” in English, for native and non-native speakers, may refer to several things.

The majority of the explanations contain “*moral principles,*” “*standards,*” or “*imperatives*” that “*govern*” or “*guide*” our “*behaviour,*” “*action,*” or “*conduct.*” According to these answers, moral principles concern the distinction between “*right and wrong*” or “*good and bad behaviour.*” This set of answers can be considered colloquial. In everyday language “ethics” may indeed mean exactly or nearly the same as morality; and many participants apparently interpreted “ethics” in this basic manner.

Only three definitions explain ethics clearly as the branch of knowledge (moral philosophy) that deals with moral principles. The answers that contain “*the philosophical study of moral values and systems*” and “*the science of moral*” and “*conceptual (or philosophical) framework for capturing morality (or a range of moral considerations)*” can be classified as philosophical approach.

In some answers we can identify a divide between descriptive and normative definitions. While descriptive ethics is an empirical investigation of people's moral beliefs, normative ethics investigates into questions that arise when considering how one ought to act, morally speaking. Among the definitions that are descriptive in nature, we could find "*principles that guide moral and social consensus*" and "*what is wrong in the eyes of the general population.*" An answer added to the definition the empirical element that morality is "*influenced by culture, tradition and religion.*" By contrast, a good example of normative definition can be: "*Ethics to me refers to questioning what the right thing is to do, what is responsible for doing, and why must we do it.*"

Three definitions focus on bioethics as applied ethics, examining specific moral issues rather than ethics as a general term. In all probability, the subject matter and the context of the questionnaire turned these students' attention towards "*biomedicine,*" "*medical ethics,*" and "*handle people remain healthy and safe.*"

In addition to the foregoing, in each section the students were asked to name not more than five principles or rules in hierarchical order, according to their presumed importance. They were also invited to answer the question: "*Why do you think these principles are important?*" In Sections (16) and (17) the students could identify the same principles as in Section (15), in the same or a different order, or add new ones.

It would be difficult to establish that in Sections (15)-(17) the answers could lead us to a solid conclusion or unquestionable quantitative results. The participating students were obviously not a representative subset of medical students or any other population group; therefore, their answers cannot accurately reflect the characteristics of a larger group. More importantly, even though 23 students provided some responses in Sections (15)-(17), almost half of the responses did not give sufficient information allowing to perceive and evaluate the students' views on the respective topics. There were several survey forms in which less than two (0 or 1) principles have been named per section. In many cases the reasoning ("*Why do you think ...?*") was missing, or simply gave a different principle rather than explaining the importance of the preferred principle. (For example, the selected principle is "*justice*", the full answer to the "*why?*" question is "*equality*".) It would be superfluous to speculate whether it was the lack of knowledge, a possible loss of interest, or other reasons that led to such insufficient answers and less effective results. The inconclusive answers still provide valuable lessons regarding the interpretation of ethical principles by the participants.

In Section (15) the students were requested to provide no more than five principles or rules in ethics in general, in hierarchical order according to the importance attributed to them by the student in question. Having considered the principles/rules named either in the highest position or generally in the whole section, a principle or rule clearly considered as dominant or paramount by the entire student group could not be identified. Among the most frequently named principles we find *justice*, *beneficence*, *autonomy*, and *dignity*. Taking the students' normative explanations into account, we can identify two broad groups of favoured principles. On the one hand, a significant number of students preferred *justice in general*, or *justice as fairness*, or *justice as an egalitarian* principle. On the other hand, the principles of *beneficence* and *non-maleficence* were named by a similar number of students. To put it differently, one can realize that deontological and consequentialist moral principles seem to be present in similar numbers and are attributed similar importance by the

students. Beyond justice, the abstract principles of human dignity and autonomy were also mentioned in the first place. For example, a student explained dignity as an alternative version of the Kantian categorical imperative as: “*Every human being exists not merely as a means, but instead always be regarded at the same time as an end.*” The Golden Rule is also named and defined explicitly as “*Treat other beings as you want to be treated by them.*” By contrast, variations of the utilitarian “harm principle”, as articulated by John Stuart Mill, could also be discovered among the answers. In addition to the foregoing, individual rights (e.g., *autonomy* and *freedom as rights*) and collective goals (*public/societal interests*) were also mentioned in Section (15), although in much smaller numbers.

In Section (16) the students were requested to provide up to five important principles or rules that, in their opinion, were particularly relevant in biomedicine. The students could take up the same principles as in Section (15) in the same or a different order, or could add new ones. There were fewer responses in this section, but they were as diverse – both in character and content - as those in Section (15). Nevertheless, it is quite evident that some of the students were already familiar with the traditional principles of Western bioethics and could articulate them in the field of biomedicine as their own. The four commonly accepted principles of bioethics, as formulated by Beauchamp and Childress, i.e. respect for *autonomy*, *non-maleficence*, *beneficence*, and *justice*, were present in different sequential orders but in a sufficiently great number to be worthy of attention. The concepts that were chosen by some students, such as *informed consent*, *privacy*, *honesty*, or *patient engagement (inclusivity)*, can also be seen as a concretisation of the abstract bioethical principle of autonomy. Similarly, *equality or non-discrimination of patients* can be understood in the context of the bioethical principle of justice. This phenomenon supports the premise according to which there are significant differences in the ethical education of medical students in the various EU Member States; however – as a common denominator – students are generally introduced to key ethical theories and principles that underpin medical ethics, such as autonomy, beneficence, non-maleficence, and justice.

In Section (17) the students were requested to provide up to five most important principles or rules that are particularly relevant for Big Data. The students could reiterate the same principles as in Section (15) and/or (16) in the same or a different order, or add new ones. The number and diversity of responses are comparable to that of Section (16). In this section 12 students named more than three ethical principles. Though the students were asked to provide principles that, in their interpretation, were relevant for Big Data, some of them selected principles particularly relevant for BDD. A student chose the very same principles as in Section (16) and explained his choice as follows: “*I don't think we are in need of new or different principles, rather we can apply general principles or specifically the principles of biomedical ethics.*” As a different type of misunderstanding, a small number of students named the three descriptive characteristics, known as the 3Vs: *Volume*, *Variety* and *Velocity*, or their synonyms, instead of giving normative ethical standards of Big Data. In this section three major ethical issues could be identified, based upon the relevant and adequate sets of answers: 1. The reoccurrence and mutual explanations of the concepts of *privacy*, *ownership*, *consent*, and *confidentiality* all referred to the moral *rights* of the *individuals*, *private customers*, or *patients*. Patients were regarded as the owners of their personal and sensitive (medical) data, and their informed and – in certain students’ opinion – clearly expressed consent was needed for using their data. 2. The concepts of *transparency*, *accountability*, *openness*, *trustworthiness* and their

equivalents referred to the moral *duties* of data controllers, data processors or data users, both institutions and individuals. 3. Equality, non-discrimination, and similar concepts might refer, among other issues, to the problem that big data algorithms can learn, integrate and even increase social (e.g., sexist, racist) biases. These answers reflect that the participants could, although in a rudimentary manner, identify some of the major ethico-legal issues generated by the phenomenon of Big Data – the issues of privacy, patient rights, data protection, and probable negative social consequences such as discrimination.

C. Understanding and recognition of ethical issues in the context of Biomedical Big Data based on the individual case stories submitted by the students (Sections 1-12)

In Section (1), the following ethical BBD topics were addressed by the students.

1. *No information.*
2. *(Preferential) delivery of vaccines to Israel by BioNTech, gets data in return, in accordance with an agreement.*
3. *Person whose grandfather had donated sperms is contacted by an unknown uncle who had done a genetic test and got the information on the kinship because of lack of anonymization.*
4. *Controlling information that will appear on a public database contain patients' information with a genetic disease.*
5. *A machine learning algorithm to predict some aspect of the Alzheimer's disease population may be biased.*
6. *Should a pregnant woman agree with a stem cell therapy for Trisomy 21, when there are controversial studies about whether it cures Down syndrome?*
7. *Google DeepMind collects data about patients without their informed consent.*
8. *Genetic data of disease (e.g., Huntington's) and control individuals reveals some controls have the disease.*
9. *Information from people who got a COVID test is saved in an online database that may be used for statistics and decision making about COVID.*
10. *Should the minister of health of a big city trust the red warning of Google Flu Trends, predicting a very high incidence of H1N1 in the next two weeks, and take preventive action?*
11. *A smart watch is promoted that collects data.*
12. *Companies can access the data assessed through smartwatch which the employees are required to wear, and may use these data against the employees.*
13. *In low and middle income countries, the data protection regulations are not as well established as in European countries, which might lead to misuse of data around the world on cognitive reserve.*
14. *Gene sequencing leads to information about cancer risk that is given to the patient who does not want to know it.*
15. *A laboratory wants to study sequencing data from another laboratory for the expression of another protein.*
16. *Data security breaches in public and private settings are caused by mistakes; others are a result of intentional abuse by technology providers.*
17. *Next Generation Sequencing identifies a risk for a cancer type that was not the topic of the medical examination.*

18. *A genome sequencing for one health problem yields a risk for another problem.*
19. *Biomedical testing of drugs: Problem of missing data.*
20. *Identification of independent germline variants in tumour diagnostics.*
21. *Routinely patient healthcare data are collected without patient consent.*
22. *After a seroprevalence study about COVID, a data breach has occurred erroneously. A member of the research team hides a potential further data breach.*
23. *Large COVID data can cause discrimination if misunderstood or misused and can lead to misappropriation of resources or unnecessary treatment.*
24. *Test for the expression of several predisposition genes to a specific disease (positive), instead of the scheduled paternity testing.*
25. *Therapeutic decisions are dominantly informed by an artificial intelligence algorithm with validity problems but that was granted a priori trust by the physicians.*

The great variety of the proposed topics reflects that the participants weren't only interested in the issues of Biomedical Big Data but also perceived that other recent technical developments in the field of medicine, such as genetic testing and genome sequencing, or the use of artificial intelligence for diagnostics or prediction, already generated absolutely new ethico-legal challenges for health care professionals and patients as well.

In order to select the proposal that was to be used as the dilemma story in order to comply with Task 2 (*"In groups, analyse the chosen and adapted case story: identify ethical considerations, do a basic research on the case story, identify the potential options the protagonist of the story might have and the decisions he/she might face, as well as conclude your stand or approaches to the story"*), Task 3 (*"In the same group, prepare a PowerPoint presentation to present the collaborative work, analysis and conclusions"*) and Task 4 (*"In the same group, present the PowerPoint to the DATAETHICS experts in an online session"*) with the help of a tutor who was only superficially informed about VaKE, certain criteria were applied to the cases suggested by the students.

The AVaKE *Report on the Virtual Exercise* prepared by AVaKE as an internal project's application observation (Intellectual Output 1; 15 May 2021) explains that moral dilemmas as used in VaKE must satisfy several conditions.

"A moral dilemma is a situation in which the protagonist has to choose between two or more (moral) values; whatever he or she does (which behaviour option he or she chooses), he or she will break one of the values. For VaKE, the dilemmas are conceived in such a way to trigger questions relating to knowledge. Key features are, among others (Weyringer & Pnevmatikos, in prep.):

- *The dilemma is personalized, i.e., a certain person has to take a decision among the options provided. The protagonist has to be given a name. The dilemma discussion is about what, in the opinion of the participants, the protagonist should do.*
- *The participants should be enabled to identify with the protagonist.*
- *The dilemma story must contain sufficient information to start a discussion.*
- *The dilemma story should be written in a narrative, engaging style."*

The internal DATAETHICS AVaKE *Report on the Virtual Exercise* concludes the student's contributions as follows.

“Many case descriptions were very detailed, while others had only some keywords. Due to the answering format, most had clear protagonists whose decision were at stake. However, while many of the stories were clear dilemmas, others were not so obvious.

In the final questionnaire after completion of all four tasks (see below), some students also addressed Task 1:

- The first individual exercise was challenging because you had to be creative to think of an interesting and suitable situation.*
- It was also nice that the first exercise asked us to think of five Ethics principles, which provides a good basis for the second exercise.*

(...)

One can see that the topics are quite widespread. Five main issues emerge: (1) Data collected in large-sample surveys (in some cases without informed consent) are used for other purposes than originally intended and declared. (2) Unintentional data breaches occur on different levels. (3) Information gathered through big data assessments are used for decisions that regard individual people. (4) Following participation in a big data assessment, people (possibly) get information they did not ask for and possibly do not want to get. (5) Methodological flaws of the big data assessments may be neglected. The big data collection tools addressed vary also considerably, including health data gathered by institutions of the health system (e.g., COVID); health data gathered by commercial companies; genome sequencing; artificial intelligence and machine learning; smart-watches; and testing of treatment success.

(...)

All this shows the scope of ethical issues in BBD as perceived by the students. This does not mean that all students have the full scope in their repertoire, as the choice of the dilemma is strongly influenced by the students' primary interest, otherwise they would not be able to give very specific information, as they often do, including literature references and details on the specific case. For VaKE, this means that the participants, on one hand, have specific knowledge about problem domains and the associated values before entering the VaKE process; on the other hand, they vary much with respect to the specific knowledge. Both aspects are valuable assets for the use of VaKE within DATAETHICS, as the prior competence and the heterogeneity of the particular student group can be capitalized on – this must be considered when conceiving the further activities.

The case stories can further be used as sources for conceiving dilemmas, not only for the tasks two to four, but also for further activities of DATAETHICS.”

As a dilemma story, the DATAETHICS expert team chose the Pfizer case proposed by student 2. Background information could be found in a Reuters article under the title *“Israel sharing COVID-19 data with Pfizer to help fine-tune vaccine rollout.”* (<https://www.reuters.com/article/us-health-coronavirus-pfizer-israel/israel-sharing-covid-19-data-with-pfizer-to-help-fine-tune-vaccine-rollout-idINKBN29N1NE>)

The dilemma story was formulated in a narrative style, with the protagonist identified, as follows.

“Israel’s Health Ministry made public most of a 20-page collaboration agreement it signed with BioNTech Pfizer whose goal is claimed to monitor the evolution of the epidemic over time and at different vaccination rates”. Israel will provide Pfizer with epidemiological data such as the number of

confirmed COVID-19 cases, hospitalizations, how many patients were on ventilators, how many died, as well as age and other demographic breakdowns (for basic information see Reuters: <https://www.reuters.com/article/id1NL8N2JT3FH>).

Hannah is the editor-in-chief of a very prominent high-circulation Israeli newspaper that is known to be critical of the Israeli government. She is not sure how to judge the deal of the Israeli government with BioNTech Pfizer. David, one of the newspaper's most prominent investigative journalist, did some research and approved the deal. Aaron, in contrast, a sophisticated political commentator, opposes it. In the redaction conference, a heated and controversial discussion emerges, as both positions, in favor and against the agreement, have about the same number of supporters. The decision is that both David and Aaron will write an article each and that Hannah has to write the feature article. In the tradition of the newspaper, the feature article takes a clear position, either pro or con. The redaction conference decides that this tradition should be continued. As editor-in-chief, Hannah has the choice of what stance to take. She is not sure.

What do you think: Should she write in favor of the deal or criticize it as inappropriate? Why?"

IV.3.3. Understanding and Recognising Ethical Issues in the Context of Biomedical Big Data in the Student Presentations

In nine groups, the students had the tasks of identifying ethical considerations, potential options, and the decision the protagonist might make (Task 2); preparing a PowerPoint presentation to present the results of collaborative work, the analyses and conclusions (Task 3); and presenting the PowerPoint in an online session (Task 4). In the final presentation via Zoom, divided into nine groups, 31 students participated.

The tasks were to be achieved in groups of students of the same university, who were to be accompanied by one tutor for each university. The recruiting of the tutors was up to the respective universities. In the final presentation each group was given five minutes to use as they chose, then an expert panel gave them feedback. The panel was moderated by C. Bödör and consisted of N. Dickman as an external expert (from Bar Ilan University), R. Mc Manus, A. Papantonis, J.-L. Patry, J. Puetz, G. Werner-Felmayer, and S. Weyringer.

A. Ethical considerations and argumentation

The large majority of the students were very motivated, engaged in fruitful discussion, and - according to the informal remarks of the experts - provided good final presentations. Both the PowerPoint presentations and the Zoom recording of the virtual exercise reveal that many students were uncertain initially; after the group discussions, however, fewer students appeared undecided or ambivalent, and several students had changed their respective opinion.

The combination of individual and group work, finding facts, identifying ethical issues, and prioritizing values and interests proved to be an excellent method to achieve the project's goal. In this way the students reflected on and eventually revised their beliefs about the given context of Biomedical Big Data, moral or non-moral. We can refer to this process as a variety of "reflective equilibrium"

methods carried out by the students acting separately and together. In the latter case, the method was dialogical, and agreement among participants could or could not be accompanied by a search for coherence.

The PowerPoint slides demonstrated that the majority of the student groups examined their moral judgements about the given Big Data issue by seeking to remain coherent with their personal beliefs and convictions about a broader range of factual and moral issues. It was a way of clarifying for themselves just what they ought to do as protagonists. Taking everything into account, the students gained a deeper understanding of a deliberative engagement, which can be seen as an amalgamation of descriptive analysis and normative account of justification, deliberating about what is right, and finding justification for their views. As one of the participants' feedback pointed out, *"In a diverse democratic society, it is something that we have to negotiate, and this is a constant process of further developing."* Given that – among several other goals – VaKE promotes personal growth and self-awareness, as the method also focuses on assisting students to develop their positive character traits such as empathy and their moral compass, and by reflecting on ethical values the students have the opportunity to develop a deeper understanding of themselves and their beliefs, the foregoing was a significant achievement of the virtual exercise.

B. Ethical issues

Overall, the PowerPoint/Zoom presentations of the ethical dilemma proved to be beneficial for the purposes of the project. The dilemma yielded differentiated perspectives, and the students provided several important principles or rules that are indeed relevant in biomedicine.

The participants took the moral rights and duties of different groups of individuals (e.g. patients, medical professionals, residents of Israel, members of the scientific community, people of other countries) and also the responsibilities of companies and institutions etc. (Pfizer, private and public hospitals, BioNTech investors, insurance companies, the government of Israel) into consideration.

In different ways, most groups presented an ethical balance between benefits and harms on both the individual and societal level. Among the frequently mentioned, most important ethical issues we could find autonomy and privacy of the patients and informed consent to data sharing and publication. However, the sensitivity of the data, transparency of information, the problems of anonymity and data management were not always in focus – which is comprehensible to a certain extent, because these issues are not merely ethical but also significantly legal, and the students were not trained in data protection laws and the GDPR.

The problems of morality and business interests, such as competitive advantage, using sensitive data to develop marketing strategies, economic benefits, or sharing data with only one company (monopolization) were also typical elements of the reasoning.

Some groups also referred to ethical issues of global justice in the context of BBD. For example, as it was pointed out by them, the agreement can cause a medical imbalance between countries that can afford the vaccine and those that cannot. As a result, the Pfizer case may contribute to the global divide between wealthy and low-income nations.

Politics can also be a moral problem; therefore it is also a relevant consideration that the government can use this situation as an incentive for the future elections. However, as the *AVaKE Report on the Virtual Exercise* points out, during the final session it was clear that some participants were not able to separate their conclusions from prior political biases and were addressing problems outside the scope of the suggested dilemma.

The presentations also raised the highly complex problem of the relationship between morality and law. (*"The agreement is legal, but this does not mean that it is ethical."*) Similarly, the theoretical difficulties on the subjective or objective nature of morality could also be detected in the presentations. Beyond the constant theoretical disagreements, the students' lack of theoretical and conceptual background limited the scope of reasoning in the presentations.

Most groups clearly understood the importance of empirical facts (e.g., efficiency and side effects of the vaccine, epidemiological reasons to vaccinate people) in the context of the Pfizer case story.

Since many aspects of the problem were not transparent (i.e. parts of the agreement were not published), the Pfizer case story offered a necessarily imperfect way of understanding and recognition of ethical issues in the context of BBD.

C. Decision making

The students were requested to decide individually and vote three times (prior to, during and after the group discussion): Should Hannah, the protagonist, write in favour of the deal (approval) or alternatively, against the deal (disapproval)?

As it could be seen in the PowerPoint slides and the Zoom recording of the virtual exercise meeting, the nine groups discussing the same dilemma reached quite different conclusions, and their insights with respect to the facts and the moral principles varied considerably. There was neither unanimous choice nor a clear priority for the protagonist writing in favour or against the deal. While a more significant part of the groups voted finally by a majority against the deal, one group unanimously favoured the deal. Many participants expressed that even though they had voted in favour or against, they were still uncertain whether this was the right decision. Hence, the dilemma turned out to be quite ambiguous.

In this way, a "reflective equilibrium" understood as the end-point of the deliberative process could not be reached. Several presentations demonstrated that the participants could not arrive at a degree of unanimous conclusion, but at least against some leading alternatives. In addition, every group made a significant effort to find relevant moral arguments in the context of the Pfizer case story and present a justifiable, coherent and ultimately acceptable reasoning. Therefore, the DATAETHICS experts acknowledged all of the insights as viable; every presentation was praised with reference to specific issues, and it was apparent that the positive remarks were absolutely sincere.

As the *AVaKE Report on the Virtual Exercise* also concludes, the presentations and the ensuing discussions gave a good impression of the impact of VaKE. Therefore, VaKE seems to be an adequate

framework of reference to achieve the project's goals, and the experiences with the virtual exercise permit fine-tuning it for specific audiences and topics.

On the level of the learning goals, the second to fourth tasks sensitised most of the participants to a large or very large extent for ethical issues, the relationship between empirical facts and moral values, the complexity of related issues and how tentative the information about facts can be. Maybe the role of ethical principles was underestimated by the students and should be getting more emphasis in future DATAETHICS activities.

In their contributions and feedback documents, the students emphasized that they had learned a lot on several levels: not only knowledge about BBD and the concrete case (vaccines in Israel) but also on bioethical issues (like that a legal agreement does not necessarily justify itself as ethical), getting access to information, about the complexity of such issues and the nature of ethical discussions.

IV.4. Analysis of Student Competences Based on the Case Story Presentation in the Online Pilot School (O.M. Péter)

Within the framework of the Pilot School (held online between 12th and 16th April 2021) – in addition to being offered interactive lectures on various topics associated with Biomedical Big Data – seven student groups, each consisting of four participants, were given the task to discuss a hypothetical case and summarise the outcome of their work in the form of a joint presentation. This assignment, in a manner identical to the virtual exercise, was conceived in cooperation with AVaKE and reflected the core values of the VaKE teaching method.

The student groups were designed in accordance with certain criteria previously developed and agreed upon by the panel of experts presiding over the presentation (consisting of G. Werner-Felmayer, A. Papantonis, J.-L. Patry, S. Weyringer, J. Pieters, R. Reynolds, J. Pütz and O.M.Péter), on the basis of the core experiences and conclusions of the virtual exercise. The first one was group stability, i.e. that the students be assigned to the various groups at the beginning of the program, thereby granting them the opportunity to bond, create a work schedule, discuss their ideas extensively, research the various issues identified during their discussions, and jointly develop their presentation. The second aspect to be considered was diversity, i.e. that each member of the various groups hail from a different university, the various levels of academic progress (undergraduate/master/PhD) be represented in a balanced manner, and gender equality be respected as well.

As the closing event of the Pilot School, the presentations were given on 16th April 2021.

IV.4.1. The Assignment

In a manner similar to Tasks 2 to 4 in the virtual exercise, the students, having received the same case story on the first day of the Pilot School, were expected to identify ethical considerations, then do a basic research on the case story, identify the potential options the protagonist of the story might have and the decisions he/she might face, as well as conclude their stance or approaches to

the story. The same group then was expected to prepare a PowerPoint presentation to submit their collaborative work, analyses and conclusions. As the final task, the groups were to present the PowerPoint to the DATAETHICS expert panel in an online session.

The case story developed for the assignment by a group of experts (in collaboration with the representatives of AVaKE) was based on a controversial case from the United Kingdom, investigated and ruled upon by the Information Commissioner's Office; the so-called "Google DeepMind" case. Based on the real-life facts and circumstances, the following fictitious story was presented to the students.

"Dr. Watson's dilemma

Dr. Margery Watson is a specialist in kidney diseases and surgery, and so far, she has fully concentrated on the well-being of her patients and research about this, and has not cared about health politics and large sample data collection. She has been hired by the Royal Free NHS Trust as Senior Clinical Fellow in Urology with the speciality Urology/Kidney Cancer at the Royal Free Hospital for a period of 12 months with the possibility of extension up to ten years; orally, the possibility of a subsequent tenure-track position has been mentioned. This is a wonderful opportunity for her as it permits her to follow her primary professional interest.

Proudly, she tells her friend, Peter Miller, about this opportunity. Miller, who works in the area of the Data Protection Act, replies that he sees huge problems with this job from the point of view of data protection as Royal Free NHS Foundation Trust provides data of around 1.6 million patients to Google DeepMind as part of a trial to test an alert, diagnosis and detection system for acute kidney injury. This partnership led to the development of the Stream mobile app, which notifies nurses and doctors immediately when test results show a patient at risk of becoming seriously ill, and provides all the information they need to take action. DeepMind and the Royal Free NHS Foundation initially claimed that the collaboration did not require patient consent because it does not aim to do research, but was instead used for direct patient care, for which there is 'implied patient care'.

Dr. Watson has no time to do further research on the issue as she has an upcoming meeting with her new superior, Dr. Alfred Woolfolk, to discuss details about the job. She asks Woolfolk whether she will be involved in the data collection which has been mentioned by Miller, and Woolfolk confirms that this is part of the implicit job description, but that this is no problem. But of course, full compliance is the condition for the extension of the employment beyond the 12 months. She discusses the problems with Miller, who cannot understand why she has accepted the nomination and urges her to quit even before the beginning. On the other hand, the new position is the fulfilment of a long-cherished dream. What should Dr. Margery Watson do?

- *Quit even before beginning?*
- *Keep the job and comply with all the principles to optimize the chances for further employment?"*

© AVaKE (Association for Values and Knowledge Education), VaKE.eu

IV.4.2. Execution of the Assignment

As far as any previous skills and competencies are concerned, the course application questionnaire included a question regarding the eventual proficiency of the students in data ethics: "Do you think

that your institute's curriculum sufficiently covers ethics issues in the context of data collection/use/interpretation and application?" Out of 133 applicants 78.1% (104 individuals) gave a negative answer, 19.5% (26 individuals) answered that the issues in question were sufficiently covered, and 2.4% (three individuals) were uncertain in their opinions. Based on such preliminary data it was apparent that the students admitted to the Pilot School were motivated and quite eager to become familiar with the various aspects of health data collection, use, interpretation, and application.

Such attitude was also reflected by the answers given by 28 successful applicants in a questionnaire surveying their pre Pilot School expectations. In the answers given to the question *"What motivated you to apply to the DATAETHICS pilot school?"*, many participants pointed out the importance of Big Data, their personal interest in this segment, and their desire to profound their knowledge of data ethics that they usually considered inadequate (*"I think it is an important topic in science, especially when talking about big data, that is not sufficiently taken into account in most cases"*; *"I was mainly motivated by the difficulties I faced in the past while handling big data, that made me wish for a proper training about big data and related issues"*; *"Opportunity to learn more about the impact of big data in biomedical sciences and think about the kind of future we're heading to."*).

As far as their expectations were concerned (*"What do you expect to take out of the DATAETHICS pilot school?"*), in addition to practical considerations such as obtaining translatable and practical skills and an opportunity to network with peers and senior colleagues from various countries, many students formulated a wish to improve their understanding of ethics, broadening their horizons and refining their skills in ethical analysis, debate and argumentation (*"I expect to have both a better understanding of ethics and how it applies to big data, as well as more nuance in my view when it comes to ethics/the morality concerning certain topics"*; *"Improve my ethical reasoning"*; *"Learning to do more ethical analytics when dealing with data"*).

While executing the assignment all groups developed a visual component; one group opted for a single poster while the remaining groups compiled PowerPoint presentations that were quite detailed and visually creative. As to the oral part of the presentations, two groups chose to give the presentation in the form of a role play where, in order to give voice to the various interests, dilemmas, and considerations, the various members of the group played the role of narrator and the characters of Dr Watson, her friend Peter Miller and her prospective superior Dr. Alfred Woolfolk.

A. Identification of certain core issues

As a first step, all groups made efforts in order to identify certain core issues, many of which appeared - formulated in various manners - in the presentations of each team. Some of the issues, however, were not purely ethical, but also (or exclusively) legal, thus demonstrating that the participants envisaged the various concepts of ethics and law as intertwined and inseparable.

Typical queries included the data types Google was collecting (*"What information is collected?"* *"Is all of the collected information necessary for the prediction?"* *"What kind of data is Google storing?"* *"What is Google doing with the data apart from using it for the application?"*); the lawfulness and ethicality of the data collection (*"Is the collection and sharing legal?"* *"What committee has approved*

the collection and to what extent?”, “Other Google DeepMind projects, are they trustworthy?”), and the position and involvement of the patients in the process, including their rights (“Are the patients informed about the app and the relation with Google?” “What do the patients know about their data being collected?” “How to protect patient data in a context with lack of transparency concerning the processing/transfer of data?”).

B. Analysis of the positions of the various parties and their respective interests

Each team identified various interested/affected parties. The parties, identified as important players and stakeholders, were Dr. Watson, the patients and the wider public (including, in three presentations, the hospital), and Google as a profit-oriented enterprise. One team analysed the probable personal motives and interests of Peter Miller and Dr. Alfred Woolfolk; another team regarded Peter Miller as a symbolic representative of the authorities responsible for data protection, and considered Dr. Alfred Woolfolk as a symbolic representative of the hospital/the NHS/the health care system.

Dr. Margery Watson

The party whose position was scrutinised in the most thorough and multifaceted manner was Dr. Margery Watson. All teams dedicated particular attention to her personal dilemma, and several teams pointed out that her decision may have an impact on a great number of patients.

The teams opined that Dr. Watson appeared to be qualified for the senior role in a prestigious and innovative university hospital. The advantages identified by the various teams included the uniqueness of the opportunity (excellent working conditions, tenure-track opportunity); that Dr. Watson has the opportunity to be part of something revolutionary in modern digital healthcare, and she can carry out her life mission as a physician, at the same time she can take part in scientific research activities and publish her results.

At the same time each team identified Watson’s main dilemma as follows: *“Does Dr Watson want to be part of a role that potentially does not value patient privacy? How would that align with her interest in public health?”* The participants adequately perceived that in such complex situations it is rather difficult to determine how to best honour the principles of beneficence and non-maleficence, and comply with the moral obligation of altruism imposed upon health care professionals by modern medical ethics.

The teams developed several arguments and counterarguments as to how Dr. Watson should approach this controversial situation. Among them, many arguments appeared “optimistic” in the sense that the teams emphasised: accepting the job would put Dr. Watson in a position where she could initiate internal reforms in order to improve compliance with the applicable privacy and data protection rules, and she could play an active role in improving the practice of informed consent, and the transparency of the data collection. At the same time the role of the “whistleblower” was also questioned, as not all teams agreed upon the moral obligation of Dr. Watson to act as one, because *“speaking out may close the door to future opportunities”*. Some other arguments questioned the empowerment of Dr. Watson, as *“Since this project is a huge collaboration between NHS and Google*

Dr. Watson's possibilities to affect the process are probably limited and outside the scope of her position." On the other hand, several teams raised the issue of Dr. Watson's moral liability if she "walks away" or doesn't do anything in order to reform the dubious data collecting practice, because in this manner she may be complicit in placing patients at risk of data breaches and exploitation against their consent.

The affected (actual and potential) patients and the health care system

Each team identified the patients as one of the major stakeholders in the complex and controversial case, and each team discussed the advantages that the actual and potential patients may enjoy by using the application; on the other hand, they identified various ethico-legal problems that violate the patients' right to privacy and the rules of informed consent. This attitude reflects that the participants were already familiar with these concepts of utmost importance.

The summary of the various advantages emphasised by the teams is as follows: The purpose of the application is to deliver better care for patients and support doctors by getting right patient data to right clinicians at the right time. The application has shown to be successful in helping clinicians detect Acute Kidney Injury at early stages (thereby reducing the workload for doctors to a certain extent); consequently there is clinical benefit, given that clinical and health outcomes are being improved (with regard to the fact that the app also reduced response time from hours to approximately 15 minutes); real lives are being improved from the collaboration, and large scale real world data has immense potential for the future development of health care services. The application has proven to be quite effective in financial terms as well, given that it reduced the cost of care to the NHS from £11,772 to £9,761 per person.

At the same time, the participants of the Pilot School were quite sensitive to the ethical and legal issues associated with the data collection practices of Google that also caused (potential) harm to the patients, and pointed out several of such issues in a correct and empathetic manner. The most recurring points of critique raised by the students were that the patients were not properly notified that their sensitive personal data had been used. According to them patients have the right to require that their data be not included in the data package to be delivered to Google; however, patients were not directly asked for consent in the sharing of their data. The participants found it ethically and legally questionable that Google gained access to patients' personal and confidential data, and historical medical data, including medical data not related to kidney disease. Three teams were particularly thorough in their analysis of certain legal issues such as data security and privacy: In their opinion the case created a dangerous precedent for (unlawful) joint mining of medical and non-medical data, the (unlawful) sharing of such data with third parties, such as insurance companies, and they also saw a risk that health analytics might be monopolised by Google to the detriment of the public, and that Google might build additional health-related tools for economic profit. They also expressed their concerns regarding data safety, as transferring and storing data is risky and increases the chance of a security/data breach. Such observations reflected that they – in a manner similar to many ethicists, health care and data protection lawyers – perceived data protection in our digital era as one of the principal issues and greatest challenges.

The position of Google as a profit oriented enterprise

The teams found the position and the proceedings of Google precarious and they expressed several doubts in this regard. The main issues identified were the unlawfulness of the data collection, the lack of transparency on behalf of Google, and the question whether the project was patient-orientated or commercially-motivated, the principal characteristic of Google being a profit oriented multinational enterprise.

One team was particularly thorough in their analysis of the role Google played in the process. According to their findings the company claimed to have no commercial plans for Google's work with the NHS. Another issue identified by the teams was the apparent incompetence of Google, i.e. that Google is expanding into healthcare but by no means is a healthcare company, so why it was selected for the task by the NHS. A team's research pointed out that such a decision was quite probably in line with a policy recently adopted by the House of Lords, i.e. that the NHS should work with AI developers in a collaborative way to share healthcare data so that it is accessible and the possible benefits to society realized. Such approach gave proof of their ability to learn autonomously, i.e. that the members of the group were able to work independently, organize the information relevant for them and integrate it in existing concepts. All teams appreciated that the issue had since been addressed by the Information Commissioner's Office and the questionable data collection and transfer practices were terminated.

Conclusions and suggestions by the teams

The teams were divided regarding whether Dr. Watson should have accepted or refused the job opportunity. While the final conclusion of four teams was that Dr. Watson should have accepted the position – although one of the groups recommended that she should accept only if she was given reassurances that she wouldn't be forced into situations where she must break any laws, or should join the hospital only after the unethical and unlawful practices were remedied – three teams concluded that, regardless of the many advantages associated with the position, Dr. Watson should not accept the job offer.

At the same time two teams also made recommendations along the line "*What should have been done better*". Their suggestions included that the data should have been anonymised; patient consent should have been sought by the hospital/NHS in advance before sharing data with Google; the patients must have been given the possibility to opt out; a proper risk assessment should have been performed regarding the potential harms and benefits arising out of the data transfer and the development of the app; the data transfer and the subsequent research should have been given a proper ethical approval by an independent ethical review board; and health data should be classified as non-commercial goods by the lawmaker so that health data cannot become subject to any sale transaction. Such recommendations reflected their ability of critical thinking and creativity.

C. Compliance with the Methodology

The teams were instructed to develop their joint presentation in accordance with the adapted principles of the VaKE teaching method and in compliance with the rules of correct scientific referencing. In the first place, participants were requested to approach the case following the steps

of “*initial vote – reasons – research – refinement – conclusions*”. While six teams properly enlightened the panel of experts in this regard, all following the steps based on VaKE as requested, one team’s presentation was not adequately clear in this regard. All teams used scientific and non-scientific sources during their research, including scientific articles, legal documents available online, and homepages offering reliable information (such as patient rights guaranteed by the NHS) to the public.

D. Assessment by the panel of experts

The assessment process reflected the aims of VaKE (<http://blog.vake.sbg.ac.at/about-vake/>). The focal points were *applicable knowledge* (when learners understand what they have learned, and are able to apply their knowledge to new situations); *development of understanding of values and their implementation in concrete situations* (when, in the discussion of concrete problem situations that represent a values conflict, learners grapple with different perceptions of values, and in that way they reach a higher level of value understanding and learn strategies in order to be able to solve moral problems); *critical thinking* (when learners learn to discuss solution statements and perceptions with each other and also to challenge them rather than simply adopting them); *autonomous learning* (when learners can work independently and organize the information relevant for them and integrate it in existing concepts, furthermore they learn to plan and supervise their learning process and also to assess themselves); *problem solving competence* (when learners acquire the necessary skills and knowledge in order to apply the learned subject matter to any given problem, to formulate hypotheses, to weigh up the alternatives and consequences of their approach to a problem, and, in doing so they end up with a self-contained solution that is considered adequate by all participants); *ability to work in a team* (when learners are able to take responsibility for each other and others in a team in a fear-free atmosphere, by introducing their ideas into the team, and also by learning active listening, making compromises, bringing up conflicts, and ensuring that everybody in the team participates and provides solutions); and also *discussion and argumentation ability* (when learners learn to stick to the declared discussion rules and to play a constructive part in the discussion, thereby acquiring the ability to formulate convincing arguments, which are based on funded facts, analogies and samples).

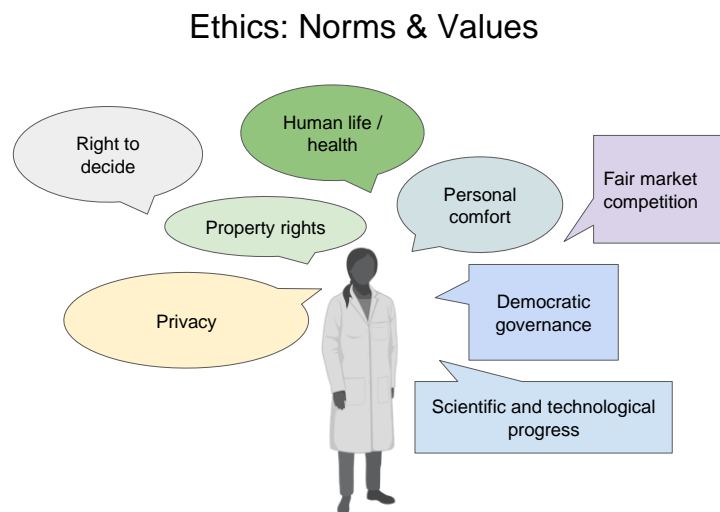
In order to assess the performance of the various groups in a uniform manner, an Evaluation Form, based on the foregoing aims and values, was created by and for the use of the members of the expert panel.

The *quality of the presentation* was assessed from the aspects whether the presentation gave only a mere summary of the case story with no interpretation or different perspective regarding DATAETHICS issues, or the presentation gave clear and well-structured insight into data ethics issues, the complexity of DATAETHICS issues of the case story was comprehensibly and creatively presented, the presentation provided food for thought and discussions to a particularly high degree. It was also assessed whether the participants failed to provide any insight about their approach and reasoning as well as the working process, or the participants clearly communicated how the members of the group approached the case and developed their presentation and cooperation.

As far as the *discussion and conclusions* were concerned, the expert panel assessed whether the conclusions were either missing or unconnected to provide insight about the case and its key elements in context to DATAETHICS, or the conclusions offered by the group members were sufficiently or wholly comprehensive so as to provide insight about the case and its key elements in context to DATAETHICS with an eventual strong creative spark.

The panel of experts considered it quite important that the participants utilise the various sources in a scientifically correct manner, therefore the *referencing methods* of the groups were also put under scrutiny. It was assessed whether references regarding sources of facts, figures and concepts were missing, inconsistent and not used according to standards, or such references were rich, correct and consistent, using an approved reference style, without any errors.

At the closing session of the Pilot School, although one of the teams was named as the best, a book prize (*"The Ethics of Medical Data Donation"*) was presented to all participants in recognition of their outstanding efforts.



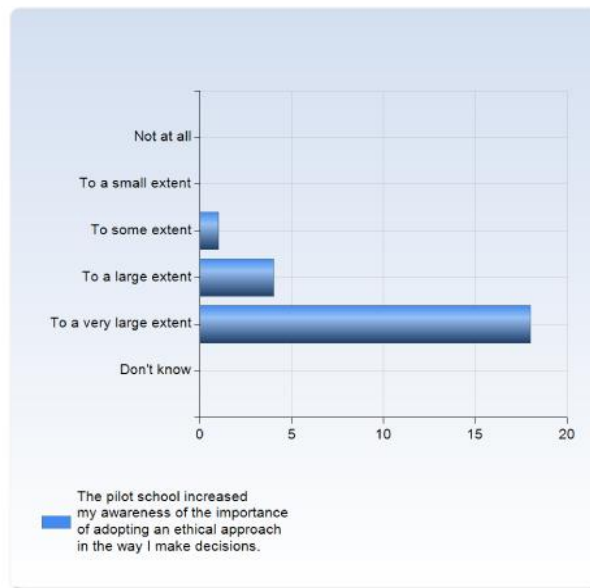
(From the presentation of Group 4)

IV.4.3. Student Feedback on Skills and Competences

After completion of the Pilot School a questionnaire was submitted to the participants, and 23 students provided insight into their experiences. Among other aspects the questionnaire included several items aimed to assess how participants opined about the possible changes in their knowledge, skills and competences that they achieved while attending the Pilot School. The students were invited to agree or disagree with certain statements by using a six-point Likert scale with six levels of agreement-disagreement.

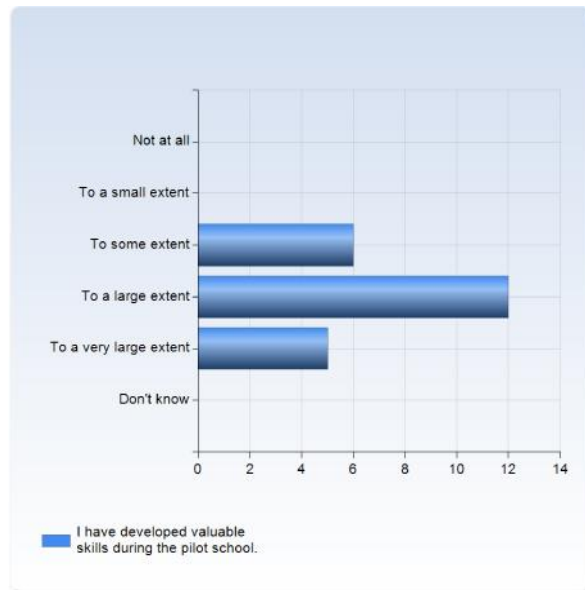
As far as their general attitude towards the importance of ethics was concerned (*"The pilot school increased my awareness of the importance of adopting an ethical approach in the way I make decisions"*), six individuals (26.1%) chose „to some extent“, 12 individuals (52.2%) opted for „to a

large extent” and five individuals (21.7%) opined that their awareness was increased „to a very large extent”.



Also, the Pilot School successfully sensitised the participants to the identification of possible ethical conundrums (*“The pilot school enabled me to identify ethical dilemmas”*); four individuals (17.4%) „to some extent”, seven individuals (30.4%) “to a large extent” and 12 individuals (52.2%) „to a very large extent”.

Among the statements concerning the general improvements experienced by the participants in their various skills (*“I have developed valuable skills during the pilot school”*), six students (26.1%) experienced such development „to some extent”, 12 participants (52.2%) „to a large extent” and five individuals (21.7%) „to a very large extent”. For some the improvement was technical (*“I felt the school supported the development of technical skills, from presenting, to panel-based discussion, to iteratively breaking down case studies in a group”*), while others appreciated the refinement of their more abstract skills (*“I gained a good theoretical understanding of the concepts involved in ethical considerations with applications to big data in biomedicine”*).



As far as certain specific skills – determined in accordance with the DATAETHICS aims and goals of VaKE – were concerned, students experienced that during the Pilot School they developed their abilities as follows.

- To think critically: „to some extent” - five individuals (21.7%), „to a large extent” - 12 individuals (52.2%), „to a very large extent” - six individuals (26.1%). Mean value: 4.0/5.0
- Work independently: „not at all” - one individual (4.3%), „to a small extent” - eight individuals (34.8%), „to some extent” - six individuals (26.1%), „to a large extent” - six individuals (26.1%), „to a very large extent” - one individual (4.3%) „don't know” - one individual (4.3%). Mean value: 2.9/5.0
- Identify relevant information: „to a small extent” - three individuals (13.0%), „to some extent” - six individuals (26.1%), „to a large extent” - seven individuals (30.4%), „to a very large extent” - seven individuals (30.4%). Mean value: 3.8/5.0
- Apply relevant information to a given task: „to some extent” - eight individuals (36.4%), „to a large extent” - nine individuals (40.9%), „to a very large extent” - five individuals (22.7%). Mean value: 3.9/5.0
- Apply different perceptions of values when discussing concrete problem situations: „to some extent” (two individuals (8.7%), „to a large extent” (13 individuals (52.2%), „to a very large extent” (nine individuals (39.1%). Mean value: 4.3/5.0
- Critically review information: „to a small extent” (two individuals (8.7%), „to some extent” (six individuals (26.1%), „to a large extent” (six individuals (26.1%), „to a very large extent” (nine individuals (39.1%). Mean value: 4.0/5.0
- Apply an ethical approach in any given issue: „to some extent” - five individuals (21.7%). „to a large extent” - eight individuals (34.8%), „to a very large extent” - nine individuals (39.1%), „don't know” - one individual (4.3%). Mean value: 4.2/5.0
- Work in a team: „to some extent” - four individuals (17.4%), „to a large extent” – nine individuals (39.1%), „to a very large extent” – nine individuals (39.1%), „don't know” – one individual (4.3%). Mean value: 4.2/5.0

- Formulate convincing arguments based on facts: „to some extent” – five individuals (21.7%), „to a large extent” – eight individuals (34.8%), „to a very large extent” – eight individuals (34.8%), „don’t know” – two individuals (8.7%). Mean value: 4.1/5.0
- Convey information: „not at all” – one individual (4.3%), „to a small extent” – one individual (4.3%), „to some extent” – four individuals (17.4%), „to a large extent” – 10 individuals (43.5%), „to a very large extent” – six individuals (26.1%), „don’t know” – one individual (4.3%). Mean value: 3.9/5.0
- Give a presentation in an effective way: „to a small extent” – two individuals (8.7%), „to some extent” – five individuals (21.7%), „to a large extent” – eight individuals (34.8%), „to a very large extent” – seven individuals (30.4%), „don’t know” – one individual (4.3%). Mean value: 3.9/5.0

The greatest improvement, as perceived by the participants, occurred to the ability of the participants to apply different perceptions of values when discussing concrete problem situations, while the least improvement was experienced in regards to their ability to work independently.

When interviewed about their level of satisfaction („Has the DATAETHICS pilot school met your expectations and in what way?”), the participants generally gave a very positive feedback: they particularly appreciated the novelty of information provided by an interdisciplinary team of expert lecturers (“School met my expectations in providing me with understanding and awareness of a broad range of topics involved in ethics, specifically in biomedical data”; „The most important were lectures and practicals with people from humanities”), the availability of experts for discussion, and the opportunity to discuss ethical issues with a group of peers as well.

When asked about the key learning points that they would take with them from the DATAETHICS Pilot School, several participants shared their realisation concerning the importance and complexity of the phenomenon which is Biomedical Big Data and the importance of interdisciplinary approach (“Biomedical data challenges lie in multiple dimensions: personal, social, legal, technical... all of these must be addressed with attention”), the importance and non-binary characteristic of ethics, the ability to consider and appreciate the various values and priorities that are different from one’s personal preferences, the importance of openness and the willingness to learn new perspectives.

What are the key learning points you take with you from the DATAETHICS pilot school?

What are the key learning points you take with you from the DATAETHICS pilot school?

The key learning point are that you are able to discuss all your thoughts with the speaker and also that you develop a group project which enhances the discussion with completely different points of view.

That there is always a responsibility that is linked to your actions, and that ignoring a problem does not mean it is not there/will stop existing (not only applicable for data ethics of course, but an important point).

The complexity of ethics, and last the differences between what legal advisors vs patients prefer when it comes to ethics + informed consent.

Some of the assumptions I have about ethical standards are not agreed upon by everyone and I need to discuss and hear more from different opinions and to always be open to learning new information and new perspectives.

1. Biomedical data challenges lie in multiple dimensions: personal, social, legal, technical... all of these must be addressed with attention.

2. Individual expertise is limited and biased - Biomedical data requires interdisciplinary collaboration.

3. Opinions must be communicated - even among the school participants there was a lot of heterogeneity regarding their values and priorities.

4. There are no universal guidelines - cases should be examined individually.

Identify ethical dilemmas in data science and biomedicine

There are no easy answers

I appreciated the most the talk about Ethics of Learning Healthcare System, of which I did not know anything at all before.

The very most important point that I learned from this school is that Dataethics is a major issue that we as responsible human beings should be well aware of. I got to understand the different perspective and rationale behind some common situations related to ethics.

Each biomedical informations have to be share/use in a reglemented context for a better protection of patient's rights and research use.

Coupling ethical and scientific considerations of the use of Biomedical Big Data, and considering different models of patient consent.

It is hard to discuss ethical dilemmas as black and white. We have to consider the shades of grey and come up with a consensus.

If the problem is complicated, don't work alone! Form a team and solve it together!

To not be so binary when considering ethical issues

To consider everything including patient care and biomedical research from an ethical point of view as well.

I learnt a lot about anonymisation and pseudo-anonymisation and the extent to which data can really be anonymised. I also learnt a lot about consent models, which I wasn't aware of that there are several models for consent.

I feel more aware of key concepts in ethics in particular the importance of patient consent, the presence of bias in different studies and gained understanding about the interpretation of results reported from big data studies. These are just a few examples that came into my mind, overall I feel more confident in discussing various aspects of ethical issues.

Complications around data ethics issues; personal values at stake, science advancing policy

the awareness of treatment of big data.

V. Conclusions and Recommendations

The interest generated by the DATAETHICS initiative among students in the partner institutions evidenced and proved the relevance of the project. It is quite obvious that – regardless of which country they hail from – a significant group of students perceive the radical and swift changes generated by a fast paced technical development in the field of medicine, and expect answers to the complex queries generated by such stellar advancements.

As far as Biomedical Big Data is concerned, students in life sciences – and medical students in particular – face multiple challenges during their education and training. Among others, students need to develop skills in managing, analyzing, and interpreting large datasets. This includes understanding data structures, statistical techniques, and data visualization tools to extract meaningful insights. Understanding and interpreting complex datasets require advanced analytical skills. Students need to develop proficiency in data analysis techniques to draw accurate conclusions.

Biomedical Big Data often involves collaboration with experts from fields such as bioinformatics, statistics, and computer science. Medical students must develop skills to collaborate and communicate effectively across disciplines and cooperate with these professionals to leverage the potential of Biomedical Big Data. Students also need to become familiar with, and avail themselves of the expertise of professionals specialised in humanities – bioethicists, lawyers, sociologist, and eventually philosophers. In addition to the foregoing, the field of Biomedical Big Data is rapidly evolving, therefore students – and other professionals – need to adapt to new data technologies, analytical methods, and data sources throughout their careers to stay current. This continuous learning experience, among others, requires skills such as critical thinking, autonomous learning, problem solving competence, the ability to work in a team, and the ability to develop understanding

of values and their implementation in concrete situations – some of the core values promoted by VaKE.

In addition to the foregoing, Biomedical Big Data often involves sensitive patient information. Medical students must learn how to handle data ethically, maintain patient privacy, and adhere to data protection regulations, and also to navigate ethical challenges related to data privacy, consent, and the responsible use of patient data for research and clinical purposes. Furthermore, Biomedical Big Data may reflect cultural and demographic biases present in the healthcare system. Medical students must be aware of such potential disparities and consider their impact on patient care.

Further to such ethico-legal considerations, students are also encouraged to engage in self-reflection and self-awareness activities to better understand their own values, biases, and potential conflicts of interest that could impact their ethical decision-making. These aspects are particularly important, given the lack of any uniform curriculum for the formation of medical students in the European Union, including the lack of any uniform curriculum for the teaching of bioethics/medical ethics. Although, apparently, we can identify certain common points in this regard – e.g. traditional core issues such as euthanasia and abortion are customarily included in the curriculum –, variations can be significant. Such differences were eminently exposed when students from various universities were requested to cooperate as project teams within the framework of the DATAETHICS Pilot School. Projects such as DATAETHICS can and indeed will facilitate the unification of learning criteria, and a rapprochement in the different methods and practices applied by the various institutions of higher education, in regards to the teaching of bioethics/medical ethics.

VaKE places an emphasis on ethical discussions and encourages participants to consider the complex moral implications of a given subject matter. By examining the ethical dimensions, participants gain a more comprehensive understanding of the topic – as supported by the experience of the students of the DATAETHICS Virtual Exercise and Pilot School. VaKE also integrates discussions and activities related to ethical values such as integrity, empathy, respect, and responsibility into academic subjects. This approach helps students understand the ethical implications of what they learn – as also supported by the testimonies of the participants of the Virtual Exercise and particularly the Pilot School. It is therefore strongly recommended that the DATAETHICS Consortium continues to apply the principles of VaKE method on the example of incorporation in DATAETHICS' early activities in its future training programs.

Vake have proven to be an effective method in other disciplines. In the field of arts and humanities, literature classes successfully analyzed certain ethical dilemmas faced by characters in literary works, encouraging students to reflect on moral choices and the consequences of actions, and the method was also successfully applied in university-based teacher education. Other probable disciplines may include social sciences, e.g. in history classes students can analyze the role of values in shaping historical events and decisions, fostering critical thinking about past actions and their consequences; sociology courses can explore the ethical dimensions of social issues – including, for example, the provision of health care by the State –, encouraging students to examine the impact of societal norms and values on different communities. In the field of law and political sciences, law students can analyze the ethical dimensions of legal cases – including, for example, cases of medical malpractice –, emphasizing the role of values in shaping legal decisions and the justice system, and

political science students can examine the ethical challenges in conflict resolution, and human rights advocacy – including, for example, the issue of patient rights as human rights. The extensive application of VaKE can also contribute to the additional improvement of education in STEM and also medicine, by creating or improving interdisciplinary relations among the various fields, such as law and life sciences.

Furthermore – based particularly on the oral and written feedback received from the participants of the Pilot School –, as far as future training programs are concerned, it is also recommended that DATAETHICS broaden its scope to encompass further, similarly novel and challenging issues such as the presence of artificial intelligence (AI) and robotics in health care, the recent developments in the field of genetics, etc. Such expansion can also be justified by the fact that Biomedical Big Data is strongly intertwined with these innovations – among others, a successful AI depends on data, and Big Data is therefore can be considered as fuel for AI. AI, on the other hand, can be used in medical robotics, diagnostics, health monitoring, digital consultation and also treatment design. These synergies and interactions are becoming stronger by the day, and soon it will be practically impossible to discuss Biomedical Big Data without taking these connections into consideration. The apparent success of the first initiatives (the Virtual Exercise and the Pilot School) and the complex experiences gained by the participants – organisers, instructors and students – warrant for the constantly improving continuation and expansion of the original initiative.

ENDNOTES

¹ Glenn Cohen—Holly Fernandez Lynch—Effy Vayena et. al.: Introduction. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 1.)

² Nicolas P. Terry: Big Data and regulatory arbitrage in healthcare. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 68.)

³ Holly Fernandez Lynch: Introduction: Oversight of Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 205.)

⁴ Nicolas P. Terry: Big Data and regulatory arbitrage in healthcare. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 56-59.)

⁵ Nicolas P. Terry: Big Data and regulatory arbitrage in healthcare. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 56-57.)

⁶ Nicolas P. Terry: Big Data and regulatory arbitrage in healthcare. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 64.)

⁷ Glenn Cohen: Overcoming the downsides of Big Data. (Introduction.) In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p.72.)

⁸ Sarah E. Malanga—Jonathan D. Loe—Christopher T. Robertson—Kenneth S. Ramos: Who's left out of Big Data? In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 98-99.)

⁹ Nathan Cortez: The Internet of Things (IoT) and Health Big Data. Introduction. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 125.)

¹⁰ Dov Greenbaum: Avoiding overregulation in the Medical Internet of Things. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 133.)

-
- ¹¹ Dov Greenbaum: Avoiding overregulation in the Medical Internet of Things. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 137.)
- ¹² Tal Z. Zarsky: Correlation versus Causation in Health-Related Big Data Analysis. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 45.)
- ¹³ Barbara J. Evans: Big Data and Individual Autonomy in a Crowd. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 29.)
- ¹⁴ Brent Mittelstadt: From Individual to group privacy in biomedical Big Data. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 178-179.)
- ¹⁵ Barbara J. Evans: Big Data and Individual Autonomy in a Crowd. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 27.)
- ¹⁶ Efthimos Parasidis: The future of pharmacovigilance. Big Data and the False Claims Act. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 73-74.)
- ¹⁷ I. Glenn Cohen: Is there a duty to share healthcare data? In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 209.)
- ¹⁸ I. Glenn Cohen: Is there a duty to share healthcare data? In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 212.)
- ¹⁹ WMA Declaration of Helsinki—Ethical principles for medical research involving human subjects. (2013). Preamble.1. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/> (Accessed: 26 June, 2021.)
- ²⁰ Laura Odwazny: Societal lapses in protecting individual privacy, the Coomon Rule, and Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 223.)
- ²¹ Laura Odwazny: Societal lapses in protecting individual privacy, the Coomon Rule, and Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 226.)
- ²² Laura Odwazny: Societal lapses in protecting individual privacy, the Coomon Rule, and Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 224.)
- ²³ Laura Odwazny: Societal lapses in protecting individual privacy, the Coomon Rule, and Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 228.)
- ²⁴ Laura Odwazny: Societal lapses in protecting individual privacy, the Coomon Rule, and Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 231.)
- ²⁵ Laura Odwazny: Societal lapses in protecting individual privacy, the Coomon Rule, and Big data health research. In: Glenn Cohen—Holly Fernandez Lynch—Effy Vayena—Urs Gasser (eds.) (2018): *Big Data, Health Law, and Bioethics*. Cambridge University Press (p. 232.)
- ²⁶ Alfred Weinberger, Jan-Luc Patry, Sieglinde Weyringer (2016): Improving professional practice through practice-based research: *VaKE* (Values and Knowledge Education) in university-based teacher education. *Vocations and Learning*, 9, p. 63-84.