United Nations
Educational, Scientific and
Cultural Organization

International Bioethics
Committee (IBC)

Distribution: limited

SHS/YES/IBC-23/16/3
Paris, 7 July 2016
Original: English

PRELIMINARY DRAFT REPORT OF THE IBC ON BIG DATA AND HEALTH

Within the framework of its work programme for 2016-2017, the International Bioethics Committee of UNESCO (IBC) decided to address the topic of big data and health, including but not limited to the issues of autonomy, consent, data protection, governance, etc.

At the 22nd (Ordinary) Session of the IBC in September 2015, the Committee established a Working Group to develop an initial reflection on this topic. The IBC Working Group, using email exchanges, started preparing a text on this reflection between October 2015 and May 2016. It also met in Cologne in May 2016 to refine the structure and content of its text. Based on the work completed so far, this document contains the preliminary draft report prepared by the IBC Working Group.

As it stands, this preliminary draft report does not necessarily represent the final opinion of the IBC and it is subject to further discussion within the Committee in 2016 and 2017. This document also does not pretend to be exhaustive and does not necessarily represent the views of the Member States of UNESCO.

PRELIMINARY DRAFT REPORT OF THE IBC ON BIG DATA AND HEALTH

PRELIMINARY DRAFT REPORT OF THE IBC ON BIG DATA AND HEALTH

I. SCOPE AND DEFINITIONS

1. Digitalization of all kinds of data is leading to an exponentially evolving phenomenon called big data. Involved technologies using these data for a vast variety of purposes are touching and transforming every area of our life all over the world. In this report the IBC addresses issues relevant to the area of health and examines what is to be taken into account so that according to Art. 3 of the Universal Declaration on Bioethics and Human Rights human dignity, fundamental rights and fundamental freedoms will be fully respected.

2. Big data is characterized by the so-called 5 Vs:

   a. *Volume* refers to the huge amount of digitalized data. It´s growing exponentially. While three-fourths of data have been analogous in 2000, today more than 99% of all data are digital data. For 2015 there is an estimated amount of 8.6 zetabyte (1021), for 2020 it is 44 zetabyte.

   b. *Variety* hints to the fact that there are different kinds of data from diverse kinds of sources. For health care and research several sources are relevant: medical data from individual patient care, public health data, data from different insurances, research data collected by researchers, companies or individuals themselves, life-style data e.g. from health apps, data from social networks and data from commerce. These data can be classified in different ways and according to different criteria: there are e.g. personal data, anonymised data, metadata, primary and secondary data.

   c. *Velocity* means the very high speed, which can be used to collect and process data. Real-time-tracking and cloud solutions allow for comprehensive processing within seconds, even producing recommendations e.g. for behaviour, drugs or nutrition.

   d. *Veracity* refers to quality of data and if they really show what they are meant to show with regard to content and precision. The context of data plays a major role here.

   e. *Value* finally draws attention to the meaning of data for a specific question e.g. with regard to a certain disease. Here again it is important to take the context of data into account.

3. Health is a normative concept which is difficult to define. Definitions vary according to the purpose of related norms. The broad definition of the WHO, that health is the state of complete physical, mental and social well-being and not merely the absence of disease or infirmity, embraces the whole life of an individual in every respect. This definition is meant as a regulative leading idea to foster global health. In national healthcare systems a narrower approach is preffered, specifying and limiting the responsibilities of medical professionals and institutions. New words like ehealth (electronic health) and mhealth (mobile health) refer to the means of healthcare in terms of electronic devices and mobile communication -and network-technologies.

4. There is a day-to-day increasing number of apps (according to estimations there are already more than 180.000) which address health issues from fitness and nutrition to diagnosing melanoma and supporting chronic disease management. It is hardly possible to sharply delineate medical from nonmedical health apps and it is even more difficult to control and to guarantee the quality of these apps. When addressing big data and health in this report, this entails more than traditional healthcare and health research.

II. PROMISES

5. During the past 10-15 years, medicine has been undergoing a revolution that will fundamentally transform the practice of future health research and health care. Technological developments in genomics and other high-throughput "omics" techniques (epigenomics, transcriptomics, proteomics, metabolomics, microbiomics, etc.) have made molecular analysis of human samples orders of magnitude cheaper and more efficient. Determination of the genome sequence of individual patients is becoming a reality.

6. Molecular characterization of diseases has shown unexpectedly high heterogeneity of common (non-communicable) diseases and rare diseases as well as cancers. This molecular data can now be complemented with digital imaging data spanning from the microscopic level to whole body imaging and with environmental and lifestyle information collected from a large number of individuals (e.g. population or patient cohorts), from surveys or from different registries, databases and research infrastructures. All this data combined with information in the electronic health records (EHR) will in the future, such is the hope, provide a fundamentally different approach to diagnose and treat patients in a personalized way, i.e. to offer the right treatment for the right person, at the right time and at the right dosage. This vision of big data in health research and care is a comprehensive and evidence-based concept of personalized medicine, more and more called precision medicine, fitting the individual patient in roughly every respect.

7. Intertwined with this concept of big data in healthcare is the rapidly accumulating big data from use of Internet, social media, smart devices, credit and customer cards etc. Such big data is being collected, analyzed and used for profiling individuals, often without them being aware of it. The Internet of Things – devices connected through the internet – allows such devices to collect and exchange data with each other and with the external users thus making analysis of very complicated and apparently unrelated data possible. Although big data collected from non-medical and non-research sources can be quite unreliable, there is a potential to use such Big Data for medical research with few restrictions as well as for profiling.

8. Furthermore it is quite likely that convergence of personal monitoring technologies and advanced wireless communication (Internet of the Body) will continue to increase and be one of the driving forces towards more reliable big data with health relevance. In the foreseeable future it is likely that health care will rapidly move towards remote monitoring of several health related parameters and activities including recommendations for health-related behaviour.

9. There is another hope with regard to remote collecting, assessing and using data. It can contribute to an improved telemedical health care for people living in regions where the next hospital or doctor is far away or in countries with low developed health care systems. Access to quality health care hopefully will be fostered by technological means, thus contributing to global health.

10. A related future scenario is that having all this in-depth knowledge of an individual's genetic make-up and "omics"-profile as well as having access to other medically relevant information like sociodemographic details, behaviour and psychological features will gradually make it possible to determine the individual's predisposition to disease and to deliver timely and targeted advice for prevention, thus blazing the trail from therapy to prevention.

11. Against this background at least four paradigm shifts in health care will occur: a shift from disease orientation to health orientation, from therapy to prevention, from health to life counseling and from patient to customer. All of these shifts address the idea that knowledge about predispositions and health-influencing factors may lead to lifestyle changes, which may make therapy less necessary. The person seeking medical advice will not be looking for treatment, but will want to know what to do to stay healthy, and so will be more a client than a patient.

12. This of course assumes appropriate actions by all the actors who are involved in this process. Members of the medical profession will have to understand and to properly use the

data. And the patients-clients will have both to understand the advice and be willing and able to follow it. However, there is at least one human-dependent weakness in this system which may change in the future when the clients are better prepared through education to deal with the new health-related data and to adjust their behaviours in an appropriate way. This weakness is the fact that even if having understood the information it does not necessarily mean acting accordingly, and in fact, often no action whatsoever is taken.

13. For pharmaceutical companies there is the hope that big data fosters molecular understanding of a disease, so that clinical studies can be precisely stratified, thus enabling studies with smaller numbers of participants, reduced costs and extended patent span. Regulators might better understand and control study designs and benefit from improved phamacovigilance. Patients finally will potentially profit from better understanding of their health status and health improving behaviour; they will have the possibility of greater control over their data (e.g. having their EHR in their smartphone) and presumably they will live better and longer.

14. It is very difficult to foresee how fast and to what degree these changes will be implemented. Some are already available – such as sharing data from many countries to analyze DNA variants or to obtain data on rare diseases. Some procedures, such as DNA sequencing, are becoming cheaper, but healthcare is expensive, and moreover, in addition to the rapidly developing implementation of precision medicine in the richer countries, there are the problems of diseases affecting people in countries with lower health-care spending. To implement big data-driven precision medicine the medical profession will need totally new solutions for data handling and representation to translate the big data into meaningful information in actual health care settings. Only then could big data be used to improve medical interventions and health services, and prediction and prevention strategies and health policies in general.

III. BLURRING LINES

15. Having in mind the broad definition of health by the WHO and recognizing that the major part of the health status does not depend on health care but on education, life style, and environment, big data opens up the way to a holistic view on health by bringing together different kinds of data from all areas of life. Thus the line between health issues and life style issues blurs.

16. This is accompanied by the blurring of the line between the health care sector and other societal sectors. Sources of health data may be the traditional ones, such as medical records, laboratory results, census data, epidemiological surveillance data, etc. but also what is not traditionally regarded as health data sources, such as social networks or consumer data bases operated by different providers, or public data sources from non-health areas such as the Internal Revenue Service, Education Departments or Social Services Departments, to name a few. Social network and search engine providers also request and collect much information on their users, which they later process and sell to different companies which in turn use “personalized marketing strategies”, offering social network and search engine users different promotions based on their search histories or participation in online groups, for instance. This also includes health issues, so that a private data search on a personal or familial condition becomes public information in the hands of companies.

17. Furthermore the line between health care and health research blurs. It can be very helpful and even life saving to have real life data about the course of illness and the effectiveness of therapies. Clinical trials happen under certain circumstances, which differ from real life conditions and they run only for a short period of time. Big data allows for data collection and assessment in a huge amount of patients over a long period of time. However, there are major challenges to data protection and to quality of data. But in principle further insights are possible, so that data from health care and e.g. from digital patient records can immediately be used for research purposes as well.

18. Also the line between different professional disciplines will blur. At least some kinds of diseases will no longer be understood according to the affected organ, but rather according to the underlying pattern of mutations and variants in the “omics”, leading to personalized therapies. A multidisciplinary approach for health care, as is already implemented in oncology in some countries, will be responsible for a patient and probably new professional disciplines will evolve.

19. Further nontraditional actors like companies will enter the health care stage and will get closer to the patient. The border between hospital and outpatient care will increasingly crumble and the smartphone will probably turn out to be a central device for coordinating one´s own health care. This will be triggered by the shift from therapy to prevention and health promotion, so that traditional health care sectors will increasingly blur as well.

IV. LEGAL CONTEXT

20. There are no specific regulations of the phenomenon of big data in the national and international legal framework. Nevertheless, there is a complete regulation about data protection in many legal jurisdictions many of whose rules can be applicable in the area of big data though big data is a new reality in sense of quantity, analysis, and accessibility. The legal provisions about data protection contain the main principles and rules to deal with the new phenomenon of big data, so there is not a lack of regulation but of specific provisions and perhaps of new principles which are adequate to regulate new features of big data.

21. If we compare national regulations, an important distinction has to be made between Europe and the US. The European approach is based on a view that privacy is a fundamental human right and it involves top-down regulation and the imposition of across-the-board rules restricting the use of data or requiring explicit consent for that use. The United States, in contrast, employs a sectorial approach that focuses on regulating specific risks of privacy harm in particular contexts, such as health care and credit. This places fewer broad rules on the use of data, allowing industry to be more innovative in its products and services, while also sometimes leaving unregulated potential uses of information that fall between sectors (Big Data: Seizing opportunities preserving values, Executive Office of the President, the White House, US, 2014).

22. In the international legal framework, the Universal Declaration of Human Rights was adopted in 1948 by the United Nations (UN) General Assembly. It covers privacy in its article 12: ‘No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.’ In a similar sense, the European Convention on Human Rights, creates in its article 8 a right to respect for private and family life, proclaiming that ‘1. Everyone has the right to respect for his private and family life, his home and his correspondence.’ The same article adds then ‘2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.’

23. The Council of Europe also approved in 1981 the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (No. 108, 28 January 1981), whose aim was to protect the individual against abuses which may accompany the collection and processing of personal data and which seeks to regulate at the same time the transfrontier flow of personal data. Later, the Council of Europe approved an Additional Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (No. 181), regarding supervisory authorities and transborder data flows, with the aim to increase the protection of personal data and privacy by improving the original Convention of 1981 in two areas. Firstly, it provides for the setting up of national supervisory authorities responsible for ensuring compliance with laws or regulations adopted in pursuance of the

Convention, concerning personal data protection and transborder data flows. The second improvement concerns transborder data flows to third countries. Data may only be transferred if the recipient State or international organization is able to afford an adequate level of protection.

24. There is as well a relevant Recommendation of the Committee of Ministers to member states in the area of biomedicine and health care: Recommendation No. R (97) 5 on the protection of medical data (which replaces Recommendation No. R (81) 1 on regulations for automated medical databank). The Recommendation protects any information, which might give an idea of a persons' medical situation, such as for insurance purposes, for example data of his or her behaviour, sex life, lifestyle, drug consumption or alcohol or tobacco abuse. The Recommendation recognizes the increasing use of automatic processing of medical data by information systems, not only for medical care, medical research, hospital management. It affirms too that progress in medical science is dependent to a great extent on the availability of medical data on individuals.

25. The OECD developed its *Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data*, 1980 (revised in 2013). There is also a recent Report of the same Organization which highlights the need for development of appropriate legal frameworks for sharing information (Improving health sector efficiency: the role of information and communication technologies, 2010). This report stresses the need for new legal framework which allows for sharing of health related information between health-care professions within and across health care organizations, as well as across organizational and geographical boundaries. The report notes that very few countries in its remit have really addressed these challenges.

26. In the EU legal framework, it is important to consider that the EU has only a limited legal competency on health matters, which can be used mainly to promote cooperation and coordination among Member States. However, in the area of data protection there is a common regulation through Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. It sets basic rights of privacy, but the exact interpretations of the practical exercise of those rights are decided in national legislation which implements the Directive. Thus a certain level of legal certainty around health related privacy exists at the EU level, but substantial variations still remain in the fine detail of the implementation of those rights in the Member States. This context will change in 2018 because of the new regulation about data protection, considering mainly its legal nature (Regulation instead of Directive).

27. In Article 8 of the Directive, a special status is accorded to all medical and health related information and prohibits the processing of health related data unless one of four exceptions is met, a) explicit informed consent; b) data processing is in the vital interests of the patient; c) the processing is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment, or the management of health-care services and the personal data in question are processed by a health professional; or d) there is a substantial public interest in the processing.

28. In January 2012, the European Commission proposed a comprehensive reform of data protection rules in the EU. On 4 May 2016, the official texts of the Regulation and the Directive were published in the EU Official Journal in all the official languages (Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) and Directive EU 2016/680 of the European Parliament and of the Council of 27 April 2016, on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such

data), and repealing Council Framework Decision 2008/977/JHA). While the Regulation entered into force on 24 May 2016, it shall apply from 25 May 2018. The Directive entered into force on 5 May 2016 and EU Member States have to transpose it into their national law by 6 May 2018. The objective of this new set of rules is to give citizens back control over of their personal data, and to simplify the regulatory environment for business. The data protection reform is a key enabler of the Digital Single Market, which the Commission has prioritized. The reform will allow European citizens and businesses to fully benefit from the digital economy.

29. The new Regulation proclaims that the processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality. This Regulation respects all fundamental rights and observes the freedoms and principles recognized in the Charter as enshrined in the Treaties, in particular the respect for private and family life, home and communications, the protection of personal data, freedom of thought, conscience and religion, freedom of expression and information, freedom to conduct a business, the right to an effective remedy and to a fair trial, and cultural, religious and linguistic diversity.

30. The new Regulation recognizes that the objectives and principles of Directive 95/46/EC remain sound, but it has not prevented fragmentation in the implementation of data protection across the Union, legal uncertainty or a widespread public perception that there are significant risks to the protection of natural persons, in particular with regard to online activity.

31. In relation to consent, the Regulation establishes that:

> [c]onsent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement. This could include ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes or inactivity should therefore not constitute consent. Consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them. If the data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided.

and later adds that '[c]onsent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment.'

32. In the area of research, the Regulation comments that:

> [i]t is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.

33. In the area of public health, the new Regulation mentions that:

> [t]he processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject. Such processing should be subject to suitable and specific measures so as to protect the rights and freedoms of natural persons. In that context, 'public health' should be interpreted as defined in Regulation (EC) No 1338/2008 of the European Parliament

> and of the Council (1), namely all elements related to health, namely health status, including morbidity and disability, the determinants having an effect on that health status, health care needs, resources allocated to health care, the provision of, and universal access to, health care as well as health care expenditure and financing, and the causes of mortality. Such processing of data concerning health for reasons of public interest should not result in personal data being processed for other purposes by third parties such as employers or insurance and banking companies.

It further states that:

> [b]y coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovascular disease, cancer and depression. On the basis of registries, research results can be enhanced, as they draw on a larger population. Within social science, research on the basis of registries enables researchers to obtain essential knowledge about the long-term correlation of a number of social conditions such as unemployment and education with other life conditions. Research results obtained through registries provide solid, high-quality knowledge which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services. In order to facilitate scientific research, personal data can be processed for scientific research purposes, subject to appropriate conditions and safeguards set out in Union or Member State law.

34. Also, the Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medical products for human use, has an important role in this area of big data.

35. On the other hand, in order to help Member States interpret their duties under the Directive, a data protection working party composed of the representatives of the national data protection authorities has been established. Its function is to advise the European Commission on the implementation of the Directive in the Member States. It is known as the Article 29 Data Protection Working Party. In 2007 the Working Party gave guidance on the processing of personal data in electronic health records (EHR) in one of its working papers (*Protection of personal data. European Commission, 2007).* The Working Party considered that consent is not a valid basis for processing data in an electronic health record. The guidance argued that because the creation of a medical record is a necessary and unavoidable consequence of care provision, a health professional may be required to process personal data in an EHR, and thus withholding of consent may be to the patient's detriment. If withholding consent could be to a patient's detriment, then such consent would not be freely given as required in Article 8(2)(a). This guidance also comments that if the safeguards for data privacy in an EHR are well drafted, it may be legitimate to offer an opt-out system. They argued that such an opt-out system would assume that for general health information a patient has opted-in to the system unless he or she explicitly opts-out. The DPWP suggested, however, that given that an EHR will contain many different types of information, such an opt-in/ opt-out system should be incremental – thus a general opt-out might apply, but a specific opt-in would be necessary for processing especially sensitive information such as information about mental health or sexually transmitted infections. The DPWP also suggested that rules should provide that a patient can prevent a particular category of medical professional seeing a particular category of his or her data. It did not say whether such suppression of data should be visible on the face of the record, but notes the value of the use of the 'sealed envelope' technique.

36. The UN Declaration on Human Rights, European Convention on Human Rights, and the European Data Protection Directive and the new Regulation are the main binding legal instruments at international level, which address privacy. It is noteworthy that two of the three international legal texts are addressed to Europe only. The effect of this fact is the European Region has a more developed legal protection of health related data than other regions (WHO,

Legal frameworks for eHealth, Based on the findings of the second global survey on eHealth, Global Observatory for eHealth series, Volume 5, 2012, p. 27).

37. At a national level and besides the regulation implemented by the Member States of the EU, the United States of America offers also a complete legal framework through some recent Acts. There are not yet regulations that concern big data but companies undertaking big data processing operations in the area of health need to comply with data protection regulation at the federal level: the Health Insurance Portability and Accountability Act (HIPAA) and the HIPAA Privacy Rule, which protect the privacy of individually identifiable health information. HIPAA is not exclusively about health data and electronic medical record, but it includes some rules about data protection. Both regulations require appropriate safeguards to protect the privacy of personal health information, and set limits and conditions on the uses and disclosures that may be made of such information without patient authorization. The rule also gives patients rights over their health information, including rights to examine and obtain a copy of their health records, and to request corrections.

38. On the other hand, the American Recovery and Reinvention Act (ARRA) of 2009 and the Health Information Technology for Economic and Clinical Health Act (HITECH) create significant incentives for an expanded use of electronic health record and they also contain some rules about data protection, mainly about measures in relation to breaches of health information.

(Comment: Some remarks on the Privacy Shield can be added later)

V. ETHICAL CHALLENGES

V.1. Autonomy

39. Autonomy in terms of exercise of self-determination entails seven dimensions:

a. Competence of the individual concerned to access, to understand, to assess and to apply relevant information
b. Information has to be available that is relevant for the question at stake
c. There has to be a choice between different options
d. Values of the individual, his or her preferences and attitudes are taken into account in deciding and acting
e. Voluntariness is granted so that the individual can decide and act without inner or outer coercion
f. Formation of will refers to the ability of the individual to choose an aim and appropriate means to reach it
g. Action can mean a conscious doing or refraining.

40. Autonomy and responsibility as well as consent and protection of persons without the capacity to consent are addressed in Art. 5, 6 and 7 of the Universal Declaration on Bioethics and Human Rights (UDBHR).

41. The two main guarantees, which have traditionally been established to protect the rights of human subjects in the area of research in health care, are consent and the disassociation or anonymization of personal data. However, it appears that in the use of big data these two guarantees present major challenges.

42. Consent has always been recognized as an important pillar of ethical research and medical practice. In the Declaration of Helsinki, consent is enshrined as a main guarantee: 'In medical research involving competent human subjects, each potential subject must be adequately informed (…) [a]fter ensuring that the potential subject has understood the information, the physician or another appropriately qualified individual must then seek the potential subject's freely-given informed consent.' Also, the Universal Declaration on Bioethics

and Human Rights of UNESCO proclaimed that '[s]cientific research should only be carried out with the prior, free, express and informed consent of the person concerned' (Art. 6).

43. It is the application of the principle of respect, which guarantees the subject's autonomy. In the context of single research projects such as in the case of clinical trials, consent has traditionally been specific, that is, subjects are asked to consent to a specific type of research. In many such cases, the consent obtained from research participants does not extend to the use of samples or data outside the original or primary research they have consented to. Valid consent requires that adequate information is disclosed and that potential subjects understand the nature, risks and potential benefits of the research and thus grant voluntary informed consent.

44. Research in the area of big data usually is not as specific as it is in other areas as clinical trials or biomedicine. The potential use of data in the future cannot be anticipated. It comes in different forms and also involves interrelationships between multiple and changing data sources, both medical and non-medical. This presents challenges to the traditional notion of consent and has led to calls for an appropriate model of consent that allows a wide use of data at the same time respecting the participant's or patient's autonomy and right to self-determination. Any model of consent for the use of big data should be tailored to the context, be it research or healthcare setting, noting that there is likely to be a huge gap in understanding between data providers and data users which could have implications for the consent process.

45. Already the field of biobanks has put forward new suggestions for consent provision such as what is called *broad consent.* Broad consent is roughly defined as consent for future unspecified use of data. It is considered particularly helpful in situations where at the time of the initial consent, it is impossible to predict how data will be used in the future. However, broad consent also raises concerns, given that the individual cannot control future uses of his or her data or, at least, this control could be weakened.

46. Broad consent is not necessarily the opposite of 'specific' consent since it may be both broad and specific in a way. It can cover a wide range of activities for a specified purpose, such as research into the causes of complex diseases, while at the same time being narrowed down to limited, though just vaguely specified uses. Broad consent typically operates at a higher level of abstraction in contrast to more narrow consent, which has a clearly defined method and aim in sight (Nuffield Council).

47. Broad consent leaves the essential informed consent model intact, but an individual simply consents to *all* possible research that could be done with her/his information in relation to a specific area or line of investigation. The individual doesn't accept through this consent all kinds of research but different in the same or similar area or line. This formula of informed consent is most prominent in the biobank context, and is utilized in Europe, where individuals broadly consent to all possible research with their tissue related to the same or similar area or line of research. It is also a model that is gaining attention in developing countries, particularly in the context of genomics and biobanking in Africa. In any case, broad doesn't mean blanket.

48. Another consent model is the implementation of opt-out consent where the research will be able to use health data unless an individual affirmatively opts out. This approach prioritizes distinct values in comparison to the opt-in formula. This prioritizes protecting informed, autonomous choices, while opt out prioritizes the collection of data. If implemented improperly, opt-out consent raises serious ethical concerns. Without adequate information about the potential implications of permitting biospecimens and data to be collected, providing the right to opt-out cannot justifiably be considered informed consent. In many cases, however, individuals are inadequately informed about the consequences given that the result of educating and informing potential donors is that fewer donors are willing to participate; practitioners therefore have little incentive to provide adequate education and information. With opt-in consent, the incentives are reversed. Because data cannot be collected for subsequent use without consent, practitioners have incentives to explain the subsequent use, even if only in general terms (Elizabeth R. Pike, 2015).

49. Irreversible or reversible disassociation of data (anonymization) does no longer appear to provide sufficient guarantee, given the interrelation and interconnection of so much data from different sources referring to one person, which makes it somewhat easier to identify the individual. The principle of anonymization as a guarantee of the confidentiality of the data no longer exists from the moment that the interconnection of data allows for the identity of the subject of the data to be derived.

50. However, though it cannot be overlooked, that autonomy and confidentiality link directly with a person's dignity, there are also major collective interests which support research being carried out in the field of big data. Therefore, a balance between individual and collective interests must be sought.

51. Even the potential benefits of public access to health information may be considerable. In an era of diminishing government funding for research, the widespread availability of high-quality datasets at little or no cost may be very important to continued scientific advancement (Sharona Hoffman, 2014).

52. In order to overcome these difficulties, a new model based on the principles of participation and transparency has been proposed. It would be a new framework facing the problems presented by technology by taking advantage of aspects presented by technology itself in order to promote greater subject participation and greater transparency from public and private institutions.

53. This new model, called 'dynamic consent', requires the involvement of public powers both as a guarantee of the individual rights as well as the promotion of individuals' participation through education and information. It calls for empowering subjects and patients to be able to monitor the use of their health data held within big data through participation, which remains guaranteed. Dynamic consent allows for control of data access by individuals, enabled by mechanisms such as consent portals.

54. As the Nuffield Council in the UK has pointed out, continuing participation can have the advantage of allowing participants to shape the possibilities of research through their decisions about what uses of data to permit by effectively 'voting' for those uses by consenting to them. While this model of consent could work well in developed countries with advanced technologies that allow patients to monitor how their data is being used through a database system, its implementation in less developed settings could face several challenges such as the lack of local technologies and low literacy rates that could hinder patients' comprehension in the context of healthcare and participants in research.

55. True participation, where patients make a choice about how information is used, and that choice is respected, serves at least two key bioethical goals. First, it prevents exploitation. Second, participation respects the patient, and preserves their dignity and autonomy. The individual effectively participates in the enterprise. In a way, the project becomes a shared or joint enterprise of the patient along with the researcher. With true informed consent, the individual truly ends up sharing the goal of the researcher, and so is not being used or exploited.

(Comment: This chapter's focus is mainly on health research. Other areas might be addressed as well.)

V.2. Privacy and Confidentiality

56. According to Article 9 of the UDBHR the privacy of the person concerned and the confidentiality of their personal information should be respected. Such information should not be used or disclosed for purposes other than those for which it was collected or consented to.

57. Such protection of privacy and confidentiality is facing several challenges in times of big data. Traditional data protection principles as purpose limitation, data scarcity and minimization, special protection of sensitive data, fair processing, and safeguarding the rights of the persons concerned have been widely implemented in order to protect privacy, though

there are different protection schemes and underlying concepts in different regions of the world.

58. However, big data inherently entails change and openness of purpose, limitlessness of data, intransparent processing, little protection of data in order to gain as much knowledge as possible, and intransparency for persons concerned. As Rinie van Est and his team worked out in an expert paper for the Global Summit of National Ethics/Bioethics Committees in 2016: ‘So while individuals are becoming increasingly transparent, our technological environment is becoming ever more opaque. ’ This is especially true with regard to algorithms getting more and more complicated and inapprehensible in times of artificial intelligence and deep learning.

59. Furthermore there are new developments coming to the front: by integrating large amounts of data from different kinds of sources more and more differentiated profiling becomes possible, which allows for acting on groups according to specific profiles, so that privacy of the individual is no longer adequately protected although the individual´s name is not known, but e.g. the IP address of his or her smartphone or computer. Thus group privacy has to be protected as well.

60. Privacy as a normative concept in addition has shown to be more than mere data protection. Several conceptualizations are discussed. With regard to big data the distinction of seven types by Finn et al. (2013) seems to be most helpful: privacy of the person (referring to bodily features and functions), of behaviour and action, of personal communication, of data and image, of thoughts and feelings, of location and space, and finally of association.

61. As a way to avoid misunderstandings, hereinafter we are going to use the term “privacy” in the sense of a right to respect for private life (what is more than personal data), in relation to those areas of life or those data that individuals want to keep reserved for themselves or, at least, for some specific members of their families or relationships. So, the term will not be used in a sense related to integrity and autonomy, as it is used in some legal jurisdiction, where privacy means something more, a sphere of self-determination of the individual where she or he has more than the faculty to exclude the others but to decide freely.

62. While individuals have privacy interests, they also share group interests in the wider use of data for health research. This broader public interest, which consists in securing objectives that are valued by society, may come into conflict with individual privacy. But the relationship between privacy and public interest is not simply one of opposition. The two are mutually implicated in each other: there are private interests in the achievement of common goals and a public interest in the protection of privacy that encourages cooperation. This complex relationship leads to a need to reconcile the articulation of the private within the public and the public within the private as the Nuffield Council pointed out.

63. From several surveys we know that people are aware of having little control over collection and mining of their data although a majority of them wants to have a certain degree of control. On the long run this will lead to a lack of trust. The loss of trust might well cause serious damage to future essential endeavours and projects to foster public health.

64. In order to protect privacy and the public good at the same time, it is essential that the population is able to trust in the good use and protection of their health data through a mix of approaches like law, governance, public surveillance, privacy by design and privacy by default.

65. New models of participation in data collection and mining, as outlined in the chapter on autonomy, go into the direction that subjects from whom data is obtained participate in the process in order to guarantee the aim foreseen by the use of data and the means, which preserve their confidentiality. Effective data protection management, accountability of use to both participants and society in general, as well as innovative models of data ownership and trusteeship are to be developed for the protection of privacy. Also a proper public education of participants and patients as well as the general population on the implications of the use of big data is of major importance.

(Comment: Confidentiality has to be addressed separately.)

V.3. Ownership

(Comment: To be added.)

V.4. Justice

V.4.1. Digital Gap

66. The digital gap (called digital divide as well) constitutes one of the major challenges in the process of democratizing information and communication, and thus development. Digital technology is, in fact, a major asset for development and for fighting poverty so that it is one of the objectives of the Millennium Development Goals of the United Nations regarding the strengthening of the global partnership for development. One target was to ensure that the benefits of new technologies, and especially information and communication technologies (ICT), in cooperation with the private sector were made available for everyone. This program actually increased the level of information of the populations, in particular the poorest and the most enclaved, the level of education and access to better quality goods and services. Regarding the health sector, this strategy strengthened the prevention of health problems thanks to a better health promotion, but also significantly expanded the supply of care through telemedicine in particular. Populations who get access to information and communication technologies (ICT), as well as professionals can inform themselves about health problems, about means for prevention and care, sometimes even before consulting a health worker. This is a positive development regarding people's awareness and knowledge, which is the basis of any public health action.

67. From 2000 to 2015, a net improvement in access to ICT worldwide could be observed, especially to mobile telephony and to mobile connectivity (Figure 1, ITU 2015). Smartphones are more and more affordable and widespread. The growing processing capacity helps to provide fluid services carrying masses of data to more people in every sector, including banks, retail trade, transportation, but also health and education. The proportion of the population covered by the mobile cellular network 2G increased from 58% in 2001 to 95% in 2015. The number of mobile phone subscriptions nearly rose tenfold during the same period, from 738 million in 2000 to more than 7 billion in 2015. These good performances in mobile telephony contributed to implement initiatives in order to strengthen the knowledge and skills for the public and actors in the fight against non-communicable diseases. For example, the *Be Healthy Be Mobile* program is implemented in several countries (e.g. India, Egypt, Senegal, Costa Rica, Zambia). This technology, being about text messages, is easily accessible for most developing countries benefiting from good mobile telephony coverage.

68. Internet penetration increased, from just over 6% of the global population in 2000 to 43% in 2015, and 46.4% of the households covered worldwide. Therefore, 3.2 billion people are connected to a global network of apps and contents, including contents created by users and social media. Rapid advancements in fixed and wireless broadband technologies constantly improve the type and quality of the available services. The wireless broadband overcame infrastructural issues, enabling more areas to connect to the Internet. Its level of penetration has been multiplied by four between 2010 and 2015 and reached 47%.

69. Despite all these advancements in access to ICT, huge disparities remain and tend to grow between countries, but also within countries. The analysis of Internet household coverage reveals clear differences according to the country's economic level. Thus, in 2015 Europe had coverage of 81.2%, well above the average coverage, against coverage of 10.7% for Africa. Barely a third (32%) of the population of developing countries uses the Internet, against 82% in developed countries. The contrast is even more striking in sub-Saharan Africa, where less than 21% of the population uses the Internet. Thus, we observe 48 Internet users less per 100 inhabitants in developing countries compared to developed countries. This puts the population of the least developed countries and developing countries in a situation of very low accessibility of the net content (including health), but also and more importantly regarding

related services. Consequently, although problems of isolation and accessibility to health facilities and professionals are given priority in most developing countries, technological innovations such as tele-expertise, tele-diagnosis or tele-consultation, constituting effective solutions cannot be implemented. This strengthens the delay in the development of health care systems in developing countries, accentuating the inequalities in access to health care.

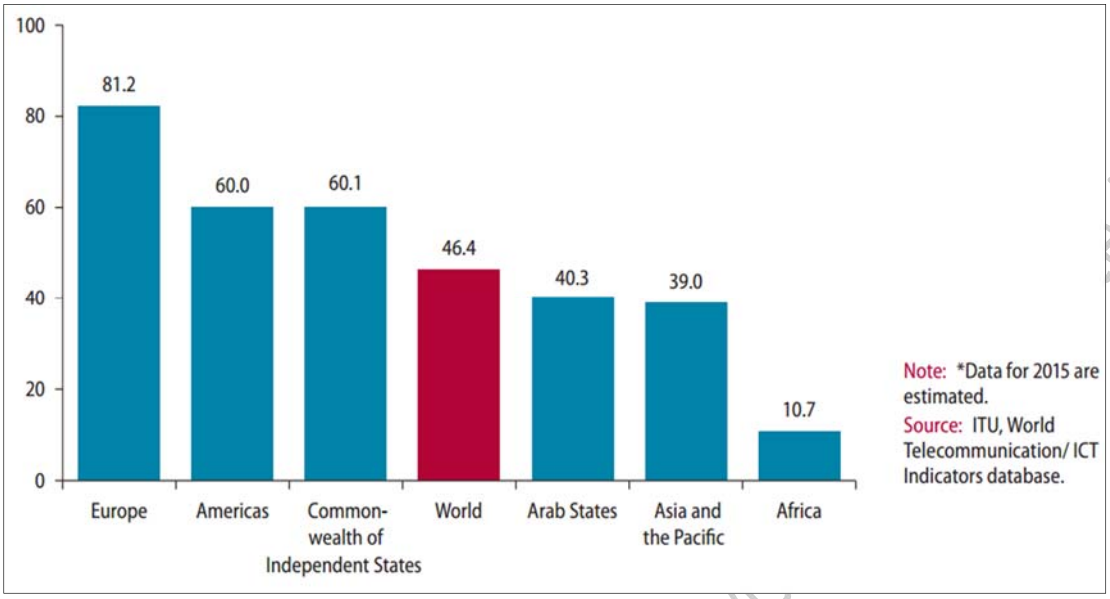

Figure 1: Covers of household Internet connection (%), 2015 (ITU)

70. Internet bandwidth and national networks' capacity are important components to provide broadband access, allowing appropriate management of big data. However, in many low-income countries, limited international bandwidth and low national infrastructure hinder the provision of broadband Internet services at affordable prices, especially in small island states and landlocked developing countries. These limitations have concrete effects on the speed and quality of Internet connections and the types of services and applications users can access. Consequently, the developing countries and the least developed countries timidly integrate the management of big data in their offers. Big data management requiring very high bandwidth with good traffic security stays little known and very poorly implemented in these countries. In addition, the average price of services remains relatively high in many of the poorest countries.

71. Despite significant decline between 2008 and 2013, the price of broadband (fixed and mobile) remains costlier in developing countries. Indeed, less than 1.6% of gross national product per capita in developed countries before 2008, it is 25.8% of per capita of gross domestic product in developing countries. In 2013, in nearly 20 countries, mainly in sub-Saharan Africa, the price of a basic Internet package of fixed broadband still represented more than 50% of Gross national income per capita. Due to the scarcity of resources in contexts of multiple priorities, developing countries are thus forced to limit their expansion in ICT. They consequently remain a major obstacle in many low-income countries, particularly in small island States or developing countries.

72. There are also significant inequalities among countries in terms of ICT skills and the existence of relevant local content. So even if the innovative content is offered through ICT in developed countries, they cannot be implemented in developing countries because it requires, apart from infrastructure, human resources for their adaptation, implementation and evaluation. This situation raises ethical problems such as massive transfer of data without clear conditions or inadequate governance policies. Indeed, masses of information are transferred from developing countries with limited human and technical resources, to be processed without some assurance on the protection of data transferred.

V.4.2. Benefit Sharing

(Comment: To be added.)

V.4.3. Non-discrimination

73. In many countries access to health care is not based on a public model but on private models, which are managed by private insurance companies with varying degrees of involvement, control and monitoring by the public powers. Big data offers the technical possibility, e.g. in the field of health care insurance, that subjects are denied or impeded from accessing insurance due to disclosure of health data. But should insurance companies become aware of the specific risks of a person in detail, it is very difficult to speak of an insurance model which has traditionally been based on balancing of collective risks rather than individual risks.

74. There are also coming up behavior-based insurance models, e.g. in life insurance, own-occupation disability insurance, automobile insurance as well as health insurance. Assured persons are offered reduced insurance rates or other kinds of premiums in case they transfer their data e.g. about their driving behavior, sportive activities, nutrition habits and the like on a regular basis via tracking, wearables and apps. With regard to health insurance, this can turn out to discriminate against three groups of people either in form of not having access to benefits or, what is to be expected in the long run, in form of suffering from disadvantages. First those people are discriminated against who don´t want to share their data with the insurance company, regardless whether they can fulfill the norms like 10,000 steps per day or not. Second those people are affected who cannot fulfill the norms because of a disability, a disease or a great misfortune, regardless of their willingness to share the data. But exactly these people are in special need of solidarity and social support. Third those people will be concerned who have different beliefs than the insurance companies about what health means. They for example prefer to support refugees, to care for a handicapped person or to engage in spiritual well-being instead of walking 10,000 steps a day. Perhaps they contribute much more to overall health, but they do not fulfill the norms given by companies.

75. Furthermore, there are powerful algorithms, which can profile individuals without having identifying data about them in order to form groups which then can get targeted advertisements, recommendations or offers. Groups can be built according to various criteria, including health-related issues. Discrimination and stigmatization then can occur against these groups, even affecting individuals that haven´t contributed data to the intitial profiling process.

76. Having this in mind, it has to be assessed with due diligence in which areas and in which ways big data can contribute to discrimination. Adequate measures then have to be taken in order to prevent discrimination and stigmatization according to Art. 11 of the UDBHR.

V.4.4. Sustainability with regard to energy and environment

77. Big data for health is also an ecological issue both for its energy cost and for the gains it can achieve through a better use of natural resources. But it is surprising to see how much the emphasis is put on the potential benefits, including energy savings and the greenhouse problem. For example, industry lobby group “The Carbon War Room” report argues that machine-to-machine (M2M) technologies can ‘facilitate 'smart grid' based efficiencies in the energy sector, optimize transportation and logistics, cut the energy footprint of buildings, and slash greenhouse gas emissions in the agriculture sector.’ ‘Smart Meters, Smart Grid, Smart Cities (…) will contribute to improving knowledge of energy expenditure and a reduction in consumption.’ But present realities are less pleasant, e.g. the French CNRS Ecoinfo service group showed how energy-intensive information technologies are and that they produce greenhouse gases at all stages of their life cycle.

78. There are infrastructures behind a distributed and remote computing that have a high energy consumption and carbon footprint. This cost is related to the operations that produce or collect, analyze and visualize digital data in mass - particularly with the joint arrival of the

Internet of Things. Giant infrastructures are already required to store this data flood, and we must add the energy cost to treat them. This inevitably has repercussions on the climate, even if the carbon impact depends on the energy mix of the respective country.

79. In use, the essential elements of big data can be divided into three categories: end devices, networks and data centers that consume each comparable electrical power. The gigantic data centers are probably the most striking aspect of this energy expenditure. They bring together thousands of machines running constantly. In addition to the energy required to operate the equipment, the huge needs for cooling must be added. In 2013 U.S. data centers consumed an estimated 91 billion kilowatt-hours of electricity (that's the equivalent annual output of 34 large (500-megawatt) coal-fired power plants – enough electricity to power all the households in New York City twice over for a year) and are on-track to reach 140 billion kilowatt-hours by 2020[1].

80. The various devices connected daily (14 billion at present, the number of objects connected worldwide will reach 125 billion units in 2025) represent another cost, mostly not for their use. Everything that is connected is now always on, even in a "sleep" mode. An estimated 80% of the energy consumption of connected objects is dedicated to maintaining their network connectivity. The International Energy Agency said that the world of connected devices consumed 616 TWh in 2013[2] (the consumption of Sweden for one year). According to a provocative scenario by Cisco company one day, perhaps in 2025, only some machines will have the right to communicate according to their pair IP address for example.

81. A novel issue might be the need to set up local Cloud storage facilities. More and more governments and businesses move their data warehouses to the Cloud. The main actors of the Cloud, such as Amazon, Google or Microsoft, engaged for years in a race to capture this market.

82. There are also concerns about the possible impact on health by pollution associated with the waste created by the connected but no longer used objects. On a global scale, the United Nations University (UNU) estimates that in 2013 there will be 53 million tons of e-waste worldwide, while around 67 million tons of new electrical and electronic equipment are put on the market. The Stopping the E-waste Problem[3] (StEP) initiative, a joint effort from UN organizations, grassroots groups and industry, predicts that by 2017 the total annual volume of e-waste will have risen by a third, to 65.4 million tons.

83. There are also concerns with the ecological impact due to search and exploitation of essential component of connected objects such as rare earths. The term rare earth describes a set of 17 chemical elements with exceptional properties. It is thanks to these rare earths that colors of our computer screens are so bright, our mobile phones have touch screens and wind turbines can generate electricity. The downside is that their extraction and their transformation pollute, produce radioactive waste and distort the landscape. Indonesia and Australia have already been impacted.

V.5. Special Challenges for Health-related Big Data Research

84. Best practice guidance concerning big data in the context of research on samples collected and stored in academic biobanks have been already issued and regularly discussed through recommendations of international organizations such as BBMRI-ERIC[4]. Recent advances include committees of patients reviewing the governance of the biobank including analysis of the use of the potential commercial value of the data, and a process of follow-up information on the various research using the data, allowing the patients to be informed on the long-range and giving them a real possibility to opt-out.

[1] http://www.nrdc.org/energy/data-center-efficiency-assessment.asp
[2] https://www.iea.org/publications/freepublications/publication/MoreData_LessEnergy.pdf
[3] http://www.step-initiative.org/
[4] http://bbmri-eric.eu/

85. To help in the follow-up of the use of the data, discussion with the editor's organizations led to initiative such as the BRIF (Bioresource Research Impact Factor) initiative, and the CoBRA guideline has been developed to standardise citation of bioresources in academic literature.

86. Several specific questions occur for research such as the problem of obsolescence of the method used to raise the results that may need re-analysis starting from the primary collected biological samples. Analogously, due to context and methods it can be necessary that data analysis has to be done again starting from primary data. Furthermore, the cost for storage cannot be supported by individual research teams and must come from research institutions.

87. Research is particularly at risk for breaches in anonymization through powerful algorithms. Aggregating data opens the possibility of re-identification through cross-referencing with data concerning ethnic background, locational data, other metadata, health records or genetic data[5]. In addition, if anonymized data subjects are grouped according to geographical, socioeconomic, ethnic or other characteristics, the anonymisation of individuals matters little if outcomes affect the groups to which they belong with a risk of discrimination and stigmatisation of affected groups. Such effects impact on all members of the community, not only those who gave consent.

88. Who owns the data might also be an ethical question even in the research field. Understood in terms of 'control', ownership grounds empowerment of data subjects to prevent any kind of data manipulation, including societal pressure to constrain any unacceptable uses. Considering the possibilities of re-identification, a control allows subjects to restrict undesired uses. For biobanks, control is relevant to considering the permissibility of using research data for commercial pursuits as is made possible by allowing private and third party companies access.

89. Many differences oppose academic/non-for-profit research to private/profit-oriented research. Personal information is being sold by providers, generating enormous revenue that is not shared with the people who used their services and provided samples and data. A well-known example is the Google based genetic test company 23&me which has sold its customers' genetic and other information to drug companies, with the stated objective of developing new drugs. People whose data is being sold did agree that it could be used for research, but it is not clear that they knew it would be sold and not donated or that it would go to private pharmaceutical companies, not academia. The same company has informed that it is planning to 'donate' its tests to underprivileged minorities so that its database would be more diverse. What effect will this have on these population groups? The ethical problems regarding direct to consumer genetic tests have been dealt with in a previous report by the IBC[6]. When they are offered to vulnerable populations in order to "exploit" their information, in times of big data these problems become even more pressing.

90. Furthermore, the blurring lines between research and healthcare, and between healthcare and lifestyle, may lead some companies to increase the market by playing on basic fears or desires. An effort of education should be done to explain the differences between correlation and causation, between big data generating hypothesis and the validation by further research.

91. The ownership of the data may also be an obstacle for research, for example preventing certain forms of data-mining or meta-analysis. National laws may limit such

[5] Choudhury, S., Fishman, J. R., McGowan, M. L., & Juengst, E. T. (2014). Big data, open science and the brain: Lessons learned from genomics. Frontiers in Human Neuroscience, 8, 239. doi:10.3389/fnhum.2014.00239.

[6] UNESCO IBC. Report of the IBC on Updating Its Reflection on the Human Genome and Human Rights. Available online at: http://unesdoc.unesco.org/images/0023/002332/233258E.pdf

research, particularly if a transfer of data is involved.

92. Another problem is coming along with the request of some scientific journals to transfer all data sets used to raise results to the repository of the review in order to allow other scientists to replicate the results or to use them in meta-analysis. But this may also lead the review to take rights on the data sets and makes data-mining prohibited or too costly. Several national laws are presently under discussion to allow free access to data sets for research.

93. Understood as a 'benefit', ownership can also require data custodians to enable data subjects to benefit and access big data for personal uses, for example knowledge of their genetic data. But such accessibility is not without risks as it opens a field of misinterpretation as well known for risk factors in the genetic domain and thus needs limitations. The general consensus is that these data being collected in the framework of a research process, they do not have the "validity grade" required in the clinic and that consequently the person will have no access to her/his personal data (such as genetic sequences).

94. Furthermore data subjects must be able to exercise their access rights with reasonable effort so that the right can be exercised meaningfully. This seems quite impossible considering all the precautions taken to protect the privacy of the data and the difficulties to interpret raw data for a lay public. Big data requires advanced scientific knowledge and technical skills that need education. Another impact is related to the loss of context that may introduce a gap between the sense for a given individual versus the sense for society as a whole. As an example in cancer, the advice of eating five different kinds of vegetable and/or fruit every day results from an evaluation of a potential decrease of 5% in the risk of colon cancer, that is to say a high impact for public health whereas a non-significant change in individual risk. Datasets do not "contain the world in small" and there is a risk that the global profile becomes a representation of the profiled[7]. To interpret the meaning of big data results for a given person remains a fundamental goal of the relationship between patient and physician. But a risk of an overreliance in big data power is the recent and on-going development of a direct relationship between the managers of the datasets and the patients excluding any kind of medical intervention. Such practice does not exist in the academic field but is rapidly developing in private research by commercial companies.

95. Big data methods will reveal unexpected results with potential health relevance for a given individual, so-called incidental findings. This is raising the question about which results people should be informed about and which way should be taken to inform them. The general consensus is to anticipate these unexpected findings and ask the dataset subjects about the way they want to know or not. Regulation in the field of big data should get example on the regulatory systems experienced in the field of genetics. Concerning incidental findings in genetics, some professional associations such as the American College of Clinical Geneticists produced lists of genes and mutations that are actionable and considered that people should be informed in case of such incidental finding. Other organizations are against such lists because results are obtained out of the context of the patient-physician relationship and lack of a technical validation process. This is particularly the case for medical imaging where the research methods might be very different from the clinical ones. In big data-research it will be difficult to list incidental findings at all because of the vast amount of possible findings which might be difficult to cluster.

96. Last but not least the objectivity of big data describing social reality is often falsely represented by overtrusting them as the ultimate edge of scientific authority. The quality of datasets used may be at stake. For instance, health records are usually data written for routine clinical work without standardization and interoperability in mind and consequently may become a source of biases. This also raises the question of robustness and the need for

[7] Bowker 2014 quoted by Mittelstadt BD and Floridi L. (2016) The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts Sci Eng Ethics 22:303–341

research to have access to the raw data for checking, meta-analysis and reinterpretation including aspects of time scale and populations.

(Comment: Aspects to be added – longitudinal translational databases with an estimated number of more than 500 mio patients, real life tracking and monitoring ...)

VI. GOVERNANCE

97. Various initiatives are underway to facilitate data sharing. Examples are the work on learning health care systems of the OECD, EU projects such as Corbel and practical instruments, which facilitate data sharing such as IMI 2014, and industry initiatives. (OECD 2013, OECD 2015, Corbel, IMI 2014, PhRMA 2014) Recently, the EMA published its guidance on the release of anonymised individual patient (trial) data (EMA 2016). These developments offer great promise, but at the same time fuel concerns about respect for the privacy rights of the people of whom personal data are processed, the clinical profession's duty of confidentiality and the protection of the interests of participants, practitioners and researchers (Cohen 2013, Laurie 2015, Lea 2015).

98. Although such mechanisms of consent and individual rights are essential to the protection of individual privacy, there are at least four dangers related to overreliance on these mechanisms in the era of big data:

a. First, individuals are no longer able to keep track of or make meaningful decisions about all uses of their personal data. The scale and complexity of personal data processing make it virtually impossible for people to keep track. Even if individuals could keep track and would be informed, their decisions might be skewed by various decision-making difficulties. In the context of health research, efforts are made to improve the process of obtaining consent by the use of online portals and engaging individuals as active participants [Kaye et al, 2014]. Although these efforts are laudable, one must recognise that individuals can engage in consent procedures only selectively. 'Broad consent' models do recognise this to some extent by inviting people to agree to a wide range of data uses. Broad consent is getting more and more accepted (WMA, CIOMS), but needs to be complemented by an ethical governance system that ensures the proper use of the data.

b. Second, even if individuals would be able to exercise meaningful control over their data, this does not in itself guarantee that all their relevant interests are protected. Although providing individuals with such control respects their autonomy, it does not necessarily reduce the risk of harm (Nuffield, 2015). Nor does it set aside the moral or legal obligations of those who are given access to the personal data, such as the obligations codified in Chapter IV of the GDPR (General Data protection regulation). Irrespective of the terms of data use set by the consent given, the interests of data subjects can only reliably be secured with effective overarching governance measures.

c. Third, relying too heavily upon consent and individual rights might conflict with the objective of removing legal obstacles to (cross-border) scientific research. In data-intensive health research, data transgress international borders and it is often impossible or impracticable to obtain individual consent. In such a situation, governance measures are necessary to justify the use of sensitive personal data in health research.

d. Fourth, the current arrangements, in which online services are provided for free, but only if the individual agrees to an unlimited (commercial) use of the data by the service provider, are clearly unfair. The advantages to the providers are disproportionately big and the possibilities for individual control are virtually lacking. All these circumstances lead to the conclusion that it is crucial to develop and apply a comprehensive multi-tiered governance structure for responsible

use of data. It is also crucial to note that, while in this time and age the individual becomes more and more transparent because we know more and more about that individual, systems tend to become more and more opaque. This needs to be countered by creating more transparent governance systems. Such a governance system can follow different models, but in all cases should be a public structure, or at least a public-private partnership. This also opens up the possibility for serious patient or public engagement, which is a prerequisite.

99. The fundamental right to data protection offers a conceptual and legal basis for such overarching and alternative governance measures, which are necessary to guarantee a morally and legally acceptable use of data. The right to data protection guarantees a complex system of information governance rules, which is more comprehensive and varied than the individual rights guaranteed by the fundamental right to privacy. Consent is for instance only one of the various grounds to adhere to the principle of lawfulness. The principle of lawfulness requires that every processing of personal data needs a legitimate basis laid down by law, but does not necessarily favour consent. Moreover, the principle of lawfulness is only one of the key principles of data protection law. Its essence is the overarching system of checks and balances, regulating the personal data processing activities of a broad range of actors through a comprehensive set of principles and rules. It aims to complement the individual rights with an effective allocation of responsibilities and duties, among the key entities involved in personal data processing. Although this system of personal data protection has its own drawbacks, its guarantee as a fundamental right provides a valuable basis to discuss and act on modern challenges of information governance.

100. This governance should address regulatory and ethical issues related to data access and release, linkage and also the governance of combined datasets of the consortium. Part of such an arrangement ought to be that individuals (can) know what they can expect when they agree to their data being collected. Governance arrangements must at least include rules for ethical oversight and arrangements for data management:

101. Ethical oversight:

a. A clear statement of the purpose of the database.
b. Procedures for consent, re-contact and re-consent.
c. Arrangements for withdrawal, and a description of the extent to which withdrawal is possible.
d. Arrangements for the protection of privacy.
e. Arrangements to provide special protection in case data of vulnerable persons or groups are used.
f. Criteria for the distribution and sharing of data.
g. Assessment of requests for secondary use of data.

102. Arrangements for data management:

a. Arrangements for the storage of data including quality control and safeguards to protect privacy and confidentiality.
b. Arrangements for access to data including arrangements for data sharing, and criteria for access, including the prioritization of research and arrangements for the settlement of competition among researchers.
c. Arrangements for the duration of storage of data and the policies after the death of a participant.
d. Arrangements for how the data will be dealt with in the event of change of ownership or closure.

103. Big data research typically depends on willingness of large groups of patients and citizens, which makes their meaningful involvement not only justified but also urgent. Despite

the variety of initiatives to involve the patients and the public, a best practice for such involvement is lacking (Kaplan et al., 2013). Examples of important questions are: What is the best way to represent these stakeholders in the organization of a Big Data governance model? What are patient and public preferences about ways to protect data while at the same time making them accessible and of high quality for the right purposes? Optimal alignment between the public, patients and other stakeholders will be necessary. First of all, arrangements for communication with participants should be established in order to be transparent, promote trust and respect their autonomy. Second, arrangements should be put in place to engage participants substantially. Therefore, a governance model also needs:

a. Arrangements that enable participants to remain informed of ongoing and novel research activities.
b. Arrangements on the disclosure of aggregate and individual research results to participants.
c. Arrangements for the engagement of participants in designing and continuously adapting the governance procedures, particularly with regard to ethical oversight and communication with donors.
d. Arrangements for benefit-sharing.
e. Disclosure of commercial interest and collaboration with commercial parties.

104. The IBC recommends that an International Treaty on Data Protection is adopted by the member states. This treaty should be based on the principle of proportionality which in this case means that any governance system has to fulfil the conditions of lawfulness, adequacy, necessity and effectivity. If so requested, the IBC is willing to prepare such a treaty. Each country can then create within its own jurisdiction country specific legislation, under which an agency can be created that is tasked with the oversight of these governance systems. Such an agency would also provide for a clear point of entry for public control.

BIBLIOGRAPHY

Cohen, J. E. 2013. What privacy is for. *Harvard Law Rev*, 126: 1904-1933.

Coordinated Research Infrastructures Building Enduring Life-science Services (CORBEL). n.d. Available at: http://www.corbel-project.eu/home.html

European Medicines Agency (EMA). 2016. *External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use*. EMA, London. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2016/03/WC500202621.pdf

European Union (EU). *General Data Protection Regulation* (compromise text after trilogue). Available at: http://www.statewatch.org/news/2015/dec/eu-council-dp-reg-draft-final-compromise-15039-15.pdf

Innovative Medicines Initiative (IMI). 2014. *Code of Practice on Secondary Use of Medical Data in Scientific Research Projects*, final draft 27 August 2014. Available at: http://www.imi.europa.eu/sites/default/files/uploads/documents/CodeofPractice_SecondaryUseDRAFT.pdf

Kaplan, W., Wirtz, V. J., Mantel-Teeuwisse, A., Stokl, P., Duthey, B. and Laing, R. 2013. *Priority Medicines for Europe and the World: 2013 Update*. WHO, Geneva. Available at: http://www.who.int/medicines/areas/priority_medicines/MasterDocJune28_FINAL_Web.pdf?ua=1

Laurie, G., Ainsworth, J., Cunningham, J., Dobbs, C., Jones, K. H., Kalra, D., Lea, N. C. and Sethi, N. 2015. On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research? *International Journal of Medical Informatics*, 84: 933-940.

Lea, N.C. 2015. *Design and Development of a Knowledge Modelling Approach to Govern the Use of Electronic Health Records for Research*. Doctoral thesis, University of London.

Mostert, M., Bredenoord, A. L., Biesaart, M. C. I. H. and van Delden, J. J. M. 2015. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *European Journal of Human Genetics*, doi:10.1038/ejhg.2015.239

Organisaton for Economic Cooperation and Development (OECD). 2013. *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Protection Challenges*. OECD Publishing, Paris. Available at: http://www.oecd.org/publications/strengthening-health-information-infrastructure-for-health-care-quality-governance-9789264193505-en.htm

OECD. 2015. *Health Data Governance: Privacy, Monitoring and Research*. OECD Publishing, Paris. Available at: http://www.oecd.org/fr/publications/health-data-governance-9789264244566-en.htm

Pharmaceutical Research and Manufacturers of America (PhRMA)/European Federation of Pharmaceutical Industries and Associations (efpia). 2013. *Principles for Responsible Clinical Trial Data Sharing*. Available at: http://transparency.efpia.eu/uploads/Modules/Documents/data-sharing-prin-final.pdf

Sethi, N. and Laurie, G. 2013. Delivering proportionate governance in the era of eHealth: making linkage and privacy work together. *Med Law Int*, 13: 168–204.

United Nations General Assembly (UNGA). 1948. *Universal Declaration of Human Rights*. United Nations (UN), New York. Available at: http://www.un.org/en/universal-declaration-human-rights/index.html

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2005. *Universal Declaration on Bioethics and Human Rights*. UNESCO, Paris. Available at: http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html