HELLENIC NATIONAL BIOETHICS COMMISSION

OPINION

Big Data in Health

Translation: Vasiliki Antoniou

HELLENIC NATIONAL BIOETHICS COMMISSION

Neofytou Vamva 6, P.C. 10674 Athens Tel.: 0030 210- 88.47.700, Fax: 0030 210- 88.47.701

E-mail: secretariat@bioethics.gr, url: www.bioethics.gr

I. Introduction

The Hellenic National Bioethics Commission, at repeated meetings, dealt with the ethics issues arising from the use of big data. The term “*big data*” refers to large amounts of information, collected by public authorities, organizations and private entities that can be extensively analyzed through algorithms in order to identify correlations and new trends.

In particular, the analysis and processing of big data regarding health provide with many possibilities, including, but not limited to: (a) identification and monitoring of health and disease trends, (b) prediction and response to epidemics, (c) preventive medicine, (d) monitoring of pharmaceutical treatments, (e) diagnosis and personalized treatment, (f) biomedical and pharmaceutical research, and (g) improvement of health and medical care policies. However, the above processing must assure the respect of the personal health data, as sensitive data that enjoy special and increased protection. The issue is timely because a huge amount of health data is already collected and stored for a variety of purposes, and can be analyzed and associated with extremely high speeds.

In our country, big data can be mainly derived from the electronic prescription system, the National Health Service Organization (EOPYY) and the insurance institutions. These data refer to the entire population of the country. They are collected, stored and are under constant processing in order to rationalize state health expenditure and reorganize health and medical care policies and insurance policies. At the meeting on February 14th 2017, the Commission organized relevant hearings of I. Ioannidis, Professor of Informatics of the National and Kapodistrian University of Athens, E. Siougle and C. Latsiou, Special Scientists of the Hellenic Data Protection Authority, and E. Fotiadou, the Deputy Director of Special Applications I.DI.KA. (Electronic Governance of Social Security SA). At the meeting on March 14th 2017, the Commission organized hearings of representatives of the National Health Service Organization (EOPYY), and in particular G. Aggouris, Head of the Department of Informatics, Ch. Kanis, Head of the Department of Planning and Monitoring of

Medicine/Drug Administration, Directorate of Medicine, and the Associate Law Officer of EOPYY, M. Diacoliou. Finally, at the meeting on November 9th 2017, the Commission organized a hearing of L. Mitrou, Professor of the Aegean University regarding the issue of the new General Data Protection Regulation (GDPR) and its application in research.

II. The technology of big data and its uses

Big data are characterized by the large amount of information and the variety they involve, the high speed at which they are collected and analyzed, the increased complexity, reliability and value generated by their analysis as they bring forth new information. Therefore, processing collections of big data provides important opportunities for improving health services, since new associations with diseases can be found, which were unfeasible by processing limited data collections. For example, by analyzing large genetic data it is possible to identify new genetic biomarkers associated with human diseases.

In addition, big data allow the search and analysis of structured and unstructured information in order to achieve better coordination and management of the offered health care services. A characteristic example is the analysis of big data coming from electronic prescription systems and electronic medical records.

Additional sources of big data are scientific data bases and study registries, hospital and insurance administration records, mobile device applications such as smart phones and smart watches, biosensors, as well as non-scientific sources, such as online shopping and personal posts on social media.

Anonymizing health data is a matter of primary importance during their use and processing in order to protect the privacy of their subject. However, it is necessary to distinguish the terms "anonymized" and "pseudonymized" data. In anonymized data, all these elements that could lead to the identification of their subjects, either independently or in combination with other data, have been removed. In pseudonymized data, instead, the details necessary for the identification of persons are replaced by a value or code (or pseudonym), but the responsible person for the processing of these data has the “key” with which it is possible to identify these persons by following a reverse process.

Along with anonymization, various other methods are used to protect the confidentiality of health information, such as encryption using algorithms and noise input. However, despite all the above measures, with the so-called “data mining”, the risk of combining available data from different sources remains, and may eventually lead to the identification of individuals.

III. Ethics issues and legislation

Health data are sensitive data. According to the applicable law (Law 2472/1997), their processing is permitted exceptionally when necessary for medical prevention, diagnosis, care, as well as the protection of public health, provided that it is performed by a health professional, who furthermore is obliged to the duty of discretion. Processing is also allowed for research and scientific purposes, provided that health data are anonymized and all necessary measures are taken to protect the rights of persons.

Issues with big data collections arise mainly from changing the purpose of collection and processing, when the data are further used for other purposes ("secondary use"). Also, as they are matched, they maximize the given information, going beyond the original purpose of the processing, which is covered by the person’s consent.

The Commission identifies three basic ethics issues concerning the creation and operation of big data bases that ought to be addressed.

A) The issue of the consent of the subjects

The consent of the subjects of sensitive health data is critical, as far as it concerns collections of big data that are “secondary”, meaning that they are composed of existing smaller collections. Even if for the creation of the latter a valid consent has been obtained by the data subjects, it is not self-evident that this consent also covers “secondary” collections, future purposes and uses. Therefore, there is a need either for a new specific consent on them, or alternative ways (“general” consent, “presumed” consent), provided that the autonomy of the persons is not affected. Nowadays, the current legislation on personal data does not make provision for such alternative ways. However, the new EU General Data Protection Regulation (GDPR),[1] which will replace this legislation from May 2018, mentions the possibility of a general consent when future research purposes of collection and processing cannot be foreordained, so it is sufficient for the initial consent to be given for possible research purposes without the need for specific reference to a specific protocol (recital 33 of the preamble). Mainly, however, the Regulation permits the collection and processing of health data for purposes of public interest, without the conditions of the prior consent and/or anonymization (article 9), differentiated on this point by current legislation.

It is obvious that the collections of big data, that are being discussed, fall into this category. However, the Commission considers that maintaining the consent of the person is a crucial guarantee of self-control of sensitive health data, in any case, even if more rigorous confidentiality safeguards measures enter into force. This is because these mechanisms do not guarantee the subject's agreement with every possible future use of the data, that is primarily ensured by the consent, which can also be inferred from the behavior of the subject, in the exceptional case where the processing is intended to protect public health.

B) The issue of confidentiality

The Commission notes the increased risk of leakage during the processing of sensitive data in large collections, precisely because of their volume. Problematic is also the ability of "deep data mining" (a process that converts primary data into

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27th 2016 on the protection of individuals with regard to the processing of personal data and on the free circulation of such data and repealing Directive 95/46 / EC (General Data Protection Regulation).

information) that can lead to the disclosure of subject identity, even in anonymous data, due to the possibility of multiple correlations.

It should be emphasized that the new EU GDPR introduces the "right to be forgotten", which is the ultimate ability to protect a person against the unlawful processing of sensitive information. Based on this right, anyone can demand the complete deletion of all his/her data from a collection, recovering the personal control of his/her sensitive information. However, it should not be overlooked that the exercise of the "right to be forgotten" may deprive the person of other important rights, particularly in the field of social security or healthcare, where the “existence” of an individual “profile” is necessary. It has to do with an individual format according to which decisions can be made about the health and insurance coverage of each individual. The national legislator will need to address the issue here, with appropriate weightings respecting both the individual's autonomy and the public interest.

C) "Incidental findings"

Finally, the third ethics issue is the handling of "incidental findings" that concern the health of the person. Incidental findings during the processing of pseudonymized data (where the person’s identity can be revealed) may occur much more frequently during data processing of large data collections, and therefore, there is a need to handle these findings with a specific policy.

The Commission has dealt specifically with this issue in a previous Opinion,[2] which it refers to. It is crucial, however, to stress that the handling of incidental findings depends on the person's special consent that must be proactively informed for this purpose.

[2] Opinion of the National Bioethics Committee on "Incidental Findings in Research and Clinical Practice" (2015). Available at http://www.bioethics.gr/index.php/el/gnomes/983-2015-09-01-09-55-51

IV. Recommendations

In the light of the above-mentioned, the Commission considers it is important to take measures to ensure the rational use of big data collections in our country. In this context, consideration should be given in particular to:

- Maintaining either a general consent (identifying the area of potential future use) or anonymization, as conditions for the processing of collections of big health data within the discretion of the national legislator given by the GDPR on personal data.
- The explicit establishment of a general consent of the data subjects for any further use and processing of information in non-anonymized data, in the context of prospective studies.
- In the case of retrospective studies, for which consent of the data subjects is not possible, the use of the data is legitimate, provided that all available measures have been taken to protect confidentiality (anonymization, encryption, noise input, etc.).
- The establishment (or activation) of Research Ethics Committees in organizations that use big data for research purposes. These Committees are primarily responsible for examining the relevant requests, approving or rejecting the access to big data collections and their processing.
- The development of codes of conduct by the above Committees for the collection, access, management and processing of big data.
- The use by the data controllers of all available data protection techniques (encryption, noise/voice changing techniques, anonymization, etc.), including the prediction of handling information leakage.
- Prior information of the data subjects (if they are identified) about the policy followed by the data controller regarding the disclosure or non-disclosure of incidental findings that may be identified and may be important for the protection of their health.

Athens, November 9th 2017