2020, February 10th

AI-REGULATION.COM

Secondary Uses of Health Data: Overview of some legal issues (CPDP 2020 Highlights)

BOUSLIMANI Katia

To cite this article:

BOUSLIMANI Katia, *Secondary Uses of Health Data: Overview of some legal issues (CPDP 2020)*, February 10th, 2020, https://ai-regulation.eu/secondary-uses-of-healthdata-overview-of-some-legal-issues/

AI-Regulation.com
CHAIR LEGAL AND REGULATORY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE

Secondary uses of health data are crucial to health researchers, but carry important risks for data subjects. Here are some takeaways of a great panel organized on this issue during CPDP 2020 and will present here an overview of the main legal issues discussed there.

The 2020 edition of CPDP was particularly interesting and mainly dedicated on "Data protection and Artificial Intelligence". Several panels focused on health data and health AI, a particular important topic. Health AI can offer great benefits to society as preventing diseases, healing, or finding remedies. However, health AI carries important risks from being flawed to be the subject of sensitive data breaches.

In this paper, we will present some highlights from one of these panels entitled "Healthy AI for access, sharing and protection of sensitive data". The principal issue that the panel tried to tackle was **secondary uses of data**, i.e. **the re-use of data for another purpose that the one for which it was originally collected**. This is one of the most important issues related to health AI-based systems in so far as AI needs a lot of high-quality data to be properly trained.

Secondary uses of data for health AI include the use of data collected by the public sector (for instance data collected by social healthcare or national registers), or the use of data collected by the private sector (for instance data collected by private insurance or via digital assistants). Secondary uses of health data should occur in an adequate data governance framework as they are both high valuable but also very risky. This framework shall guarantee the quality of the data sets used to train health AI and protect individuals against any violation of their fundamental rights.

This paper will focus on the legal basis of consent (1), on the benefits of secondary uses of health data (2) and on the issue of de-identification (3).

1. SHOULD CONSENT BE THE LEGAL BASIS FOR SECONDARY USES OF HEALTH DATA?

Lots of guidelines on secondary uses of health data consider that consent should be taken into consideration for such uses. However, the international community does not seem to have reached a consensus on consent as an appropriate legal basis in the context of secondary uses of health data.

The Finnish example

For instance, Joni Komulainen (Ministry of Social Affairs and Health of Finland) explained that Finland chose not to use consent as a legal basis for secondary uses of health data. Instead, Finland referred on another legal basis provided by Article 9 of the GDPR on the processing of sensitive data:

- "personal data which are manifestly made public by the data subject" (Art. 9.2.e);
- "processing is necessary for reasons of public interest in the area of public health such as protecting against serious cross-

border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices …" (Art. 9.2.i);

- "processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes …" (Art. 9.2.j).

The Finnish Ministry of Social Affairs and Health's choice was mainly motivated by three reasons:

- First, Joni Komulainen considers that **consent is a "weak legal basis"**. Finland relies on the guidelines on consent of Article 29 Working Party to conclude that "consent is probably not the best legal basis" for secondary uses of health data. Joni Komulainen takes the example of genomic data. In that case, an individual cannot really give his consent to the sharing of his genomic data because this data does not only belong to him. Indeed, genomic data also gives information about the relatives of the individual.

- Second, according to Joni Komulainen, not choosing consent as a legal basis also has a practical interest for the **quality and the accuracy of the data gathered**.

- Third, in Finland, the gathering of personal data on the basis of a public interest is not new for Finish citizens because data banks are gathered since the 1950s mainly without consent. According to Joni Komulainen, **citizens trust public institutions** thanks to the transparency of their law-making and decision-making processes.

However, according to Elettra Ronchi (Senior Policy Analyst at OECD), this Finnish model cannot be exported because of "the generalized sense of individuals about the lack of control". Indeed, in Finland, "the regime is justified by the trust that citizens have for public authorities, which is not reflected in other situations and conditions". Effy Vayena, an academic researcher specialized on bioethics (ETH Zurich), agreed on the specificity of Finland. According to her, Finland has "another social contract". Finnish citizens trust public authorities with their data because "they trust that the benefits will come back to them".

The EU approach

On the other side, the European Union chose to refer to a different model using consent as a legal basis. In that sense, the European Commission worked on the question of secondary uses of data in several sectors. Its experts found that "in consumer-facing markets, people believe that consent should be the key issue in order to preserve the data". To demonstrate these statements, Alberto Gago Fernandez (DG Connect, European Commission) emphasized that consent is already used as a legal basis for secondary uses of data in the banking sector since the Payement Services Directive (PSD2).

Furthermore, in the health sector, consent is not so easy to get as it has to be meaningful. Effy Vayena explains that otherwise, it gives an illusion of control which causes a double harm to citizens: "first of all, they are not given a choice, second they believe they do". According to her, there are several cases of research where “consent is necessary and the right thing to do”. However, for most of the secondary uses of health data, consent cannot be meaningful. Indeed, as Elettra Ronchi emphasized: the traditional consent system is challenged “by the fact that data today is not only provided by the individual, it is data that is observed by a variety of devices and sensors”.

As we can see from the above, consent is not unanimously considered as a necessary legal basis for secondary uses of data.

2. WHO IS BENEFITING FROM SECONDARY USES OF HEALTH DATA?

A "legacy problem"

Effy Vayena addressed what she called a “legacy problem”. According to her, the guidance is focusing **upstream** on **how to get the data** and what are the conditions to access and share the data. However, she emphasized that guidance also has to focus **downstream** on **what is the purpose of the use,** what we are trying to achieve, the exact benefit of the processing and how this benefit is going to be distributed and to whom.

She argued that these questions are relevant as they are linked to the notions of trust and trustworthiness. Such purposes cannot be achieved only through the consent of the individual. They also have to be achieved by the fact that the benefit has been approved and distributed fairly. The purpose of the use “remains the critical point” even for intergovernmental sharing and access of data. For example, if health data from one sector is used by law enforcement, it could lead to certain decisions that directly affect citizens. To avoid this kind of effects, she said, governments should aim to frameand create what she called a “social license”: a moral licence or agreement between officials and citizens where the public should benefit from the process, assess or sharing of their data and being informed of their use or re-use “in a way that is also transparent and understandable to people”.

A paradoxal behavior of citizens

Concerning the benefits of secondary uses of health data, the panel also addressed the paradoxical behaviour of citizens when it comes to the sharing of data with the private sector.

While individuals might be willing to have their health data used for general public health objectives, they are most often reluctant to the involvement of the private sector in the process because they are opposed to what they consider to be a “commercialisation of data”. For example, Joni Komulainen mentioned that the Finnish law was criticized by some people because companies benefit from the secondary uses of health data. However, according to him, if you look at the big picture, “we are all benefiting” from this sharing of data with the private sector.

The individualization of benefits

The individualization of benefits reflects a situation where individuals are voluntarily sharing their data to get “monetary incentives” like money or free services. According to Joni

Komulainen, such a mechanism is not possible at the moment (because Finland's law forbids data collectors to profile an individual), but Finland will study further how to make it possible in the future. However, according to Effy Vayena, the individualization of benefits will be problematic for mainly three reasons: first because "the real value is not in the individual data sets" but "in the aggregated, in the many"; second because "it creates a culture where you are selling a part that constitutes your identity" and third because health data sharing should only focus on the public benefit, and not on the individual one.

3. THE SENSITIVE ISSUE OF DEIDENTIFICATION

Following a question by one auditor, the panel also discussed briefly the major issue of de-identification. The auditor introduced the notion of demonstrable acountability in data protection and he defined it as all "auditable safeguards that could prove that data governance had been exercised". Thus, he asked if de-identification techniques pseudonymisation could useful for demonstrable accountability.

Joni Komulainen shared the Finnish experience. When Finland was drafting the Act on secondary uses of health and social data, their initial idea was to freely provide anonymised data because "as far as the GDPR is concerned, when it's done properly using reasonable resources, it's not personal data anymore". Its Ministry studied different anonymisation technologies and "found them all flawed". Its conclusion was that even if they get better, they would be still be flawed because "if you have other data that's anonymized somewhere and enough resources – like big companies like Google or big nations like China –, it does not take much to break down anonymisation". They finally set up a "double system": first, the anonymised data is used only in a safe and secured environment and second, a national authority checks that there is no identifiable data in the publications that used this data.

Elettra Ronchi from the OECD also provided some interesting elements about demonstrable accountability by the data controller. She advised to adopt a "risk-based approach" to demonstrate "due diligence when it comes to potential risk of re-identification", especially with sensitive data.

Finally, Riccardo Masucci (Intel) mentioned that his company and other companies are looking for "privacy-preserving machine learning which seems to be a very promising area of research". He enumerated some of these areas, such as "homomorphic encryption, differential privacy, [and] federated learning".

These statements are attributable only to the author, and their publication here does not necessarily reflect the view of the other members of the AI-Regulation Chair or any partner organizations.

This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003)