HealthManagement
Promoting Management and Leadership

GDPR: Conducting “big data” research with European health data

Eugenio Mantovani
******@***vub.be

Researcher - Research Group on Law, Science, Technology & Society (LSTS)
Vrije Universiteit Brussel

Prof. Dr. Paul Quinn
******@***gmail.com

Researcher - Law Science Technology & Society (LSTS)
Vrije Universiteit Brussel

As a result of rapid transformations in information technology, medical research is becoming increasingly data intensive. Imagine a researcher- we’ll call him Walden - who is planning, in the year 2020, to carry out a big data analysis on large amount of personal health data stored in datasets and data repositories located in Europe. Walden comes from a geriatrics university center in Vancouver, Canada, and his plan is to access personal health data concerning falls and falls related injuries. His research purpose: to improve the understanding of the physiological contributors to falls and injuries risk factors through omics analysis for developing personalised intervention models.

To attain his goal, Walden needs to assemble great quantities of personal data (what we call big data): existing data on diagnoses of real falls and resulting injury (TBI, fractures of the hip, vertebrae, and wrist, scans acquired secondary to a fall, etc.); in addition, his big data analysis requires processing clinical information on patient characteristics, demographics, biomarkers (-omics), medications, and patient profiles data such as physical activity information, functional status, etc.

Walden’s plan for big data analysis poses two obstacles. The first obstacle is technical and hinges around the existence or the lack of harmonised data, interoperability standards, and of semantical alignment between different datasets. The second obstacle, discussed in this contribution, is regulatory, having to do with the respect for the fundamental rights to privacy and data protection of patients.

In short, big data analysis upsets the paradigm “consent or anonymity”, according to which the processing of medical data for research purposes requires either individual informed consent or the previous anonymisation of personal data processed for the research. As our researcher from Canada prepares his trip to Europe, he may notice two inherent difficulties Consent to processing health data, to be valid, must be, freely given, specific, unambiguous and explicit. This means that Walden should obtain consent from the patients before starting its research. Obtaining such consent may come with a significant price, both in administrative and financial terms, given that consent may need to be gathered from a large number of individuals who may not be readily accessible (or even alive).

Walden could then decide to anonymise the personal information. Once the personal data are rendered irreversibly anonymous, the data subject is not or no longer identifiable, thus no consent in required. This option, anonymisation, which is often desired in medical data research, has its shortcomings too, both in general, and in the context of big data analytics, in particular. In general, advances in modern computing, the development of sophisticated deanonymisation algorithms, the increased availability of potentially complementary data online, have restricted the cases in which it is “reasonably impossible”, “empirically implausible” or “logically impossible”, to re-identify particular individuals from a dataset of anonymised data.

In particular, according to some scholars, technically perfect anonymous data, while protecting patients’ privacy, would render the data less useful or downright useless for big data analysis. American privacy scholar Paul Ohm is direct on this: “data can be useful or perfectly anonymous but not both’ (Ohm 2010, 1704)”. According to Ohm, there is a trade-off between privacy and utility for anonymisation techniques to be made. An analysis on completely, irreversibly, anonymised database with no information but diagnoses, for instance, would greatly protect the privacy of the patients, but reduce the utility to be extracted from the big data analysis: “Small increases in utility are matched by even bigger decreases in privacy, and small increases in privacy cause large decreases in utility.”(p.1755) If Ohm is right, while formal anonymity is vulnerable to advancements in deanonymization techniques and does too little to protect one’s identity, complete anonymisation may go too far

to be useful for Walden's research purpose.

These considerations (consent is too difficult to obtain, anonymisation is vulnerable to attacks and contradicts the exigences of big data analysis in health) question the consent/anonymisation paradigm and cut in favor of other forms of regulation that may be in place when Walden comes to Europe in 2020, one year and a half after the GDPR has entered into force. While the trajectories of these other forms of regulation remain to be defined, in our views, the incoming GDPR seemingly points at one form of regulation in which both the law, i.e. privacy and data protection laws and requirements (e.g., on consent of anonymisation), and ethics, i.e., national authorities and research ethics committees, play a role in the creation of a research space for big data analysis in health care that would limit or permit access to personal medical data. The GDPR seeks to facilitate the use of personal data in scientific research (art. 9.2; Recital 52, Recital 157): it defines scientific research purpose broadly as including "for example, technological development, and demonstration, fundamental research, applied research and privately funded research" (Recital 159), and it acknowledges scientific research as a possible basis for processing personal data (Recital 52). The Regulation, however, conditions the processing of personal data on the existence of well described scientific purposes and of "appropriate safeguards" to protect the data subject's rights, freedoms, and legitimate interests (Recital 156). In the verification of the research purpose and in the adoption of the appropriate safeguards, there is a role for data protection law, which provides for harmonized conditions, e.g., data minimisation, data security, transparency, etc. (Recital 156); and there is also a role for other norms adopted at the national level by member states which can, says the Regulation, "maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health." (Recital 53)

Although the Regulation does not mention ethics committees explicitly (indeed it is a " general" data protection regulation), in our view, research ethics committees at national, regional, and local level will be decisive to determine both the boundaries of scientific research purposes as a ground for processing medical data and to decide the "appropriate safeguards" for such processing activities. Our researcher Walden must ensure, in conclusion, that his research proposal must be both compliant with the law (e.g. data protection, privacy laws or specific laws relating to the use of medical dossiers) and with the applicable ethical approaches. He must be warned, however, that the ethical approach used in a particular instance is both highly variable and contextual varying from state to state, from region to region and even between various types of institution (e.g. university, hospital or private entity). This plurality of both the ethics approaches and the potential outcomes that can arise therefrom (even within the same legal jurisdiction) raises questions about the ability of gaining approval for various forms of research in the absence of individual consent, despite the existence of such an option being clearly available in the law, to access different repositories in various member states, and about the possibility to appeal against negative decisions.

Published on : Wed, 6 Jun 2018