



From bottleneck to enabler: a new approach to regulating data-driven medical research

Evert Stamhuis¹ · Laura Evangelista² · Sebastian van der Voort³ · Anastasia M. Mele⁴ · Elisa Spiller⁵ · Esra Demir¹ · Andrea Pin⁶ · Klaus Heine¹

Received: 22 December 2022 / Accepted: 5 February 2023 / Published online: 15 February 2023
© The Author(s), under exclusive licence to Italian Association of Nuclear Medicine and Molecular Imaging 2023

New data-driven medical technologies have been developed and are under construction that hold important benefits for patient wellbeing, disease management and clinical practices [1]. This will serve the innovation necessary for maintaining the best level of clinical practice, implying that more people can benefit from those services. For such innovations, medical researchers until recently met with technical and infrastructural hurdles. However, with these hurdles diminishing, the window opens to another bottleneck: the data and research regulations [2]. Quite frequently, the ethical and legal risks of data use stand in the way of research aimed at improving clinical practice.

Of the manifold data-driven research practices where the legal/ethical bottleneck is currently experienced, we set the spotlight on two exemplary cases. The first example is the use and sharing of magnetic resonance imaging (MRI) scan data for improving the diagnosis and treatment of brain tumors, assisted by algorithms, where the issue arises out of the higher level of safeguards that the data protection regulations postulate for processing personal health data. Researchers need large and combined datasets, especially for continuous training of the algorithms, which incentivizes data availability across institutions, sectors and even across

national borders.¹ Additionally, this example also elucidates the issue of personal versus anonymous data, but we will come back to that in more detail below. The second example regards the release of an algorithm, trained on MRI data and recognizing certain disease markers, for use by other parties outside the original institution. This does not pertain to sharing data, but that is only true with regard to the current status of the technology. In short, this case demonstrates the privacy concerns with a view to future technological developments and the competitiveness of an algorithm. Both examples point to the divergence in laws and policies across institutions, countries and continents.

Let us just demonstrate the regulatory environment for researchers by briefly looking at two trans-institutional issues. Our first example relates to data-sharing practices across institutions and countries. First of all, there is the

¹ Whether this data availability is technologically shaped as data pooling or as federated learning is not decisive for our argument here. The draft regulation of the European Commission for one European Health Data Space (EHDS) categorizes this all under secondary use of data. See: European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space', COM(2022)197/final, 03.05.2022, Strasbourg art. 34(1)(g) and (h) (https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_en).

✉ Evert Stamhuis
stamhuis@law.eur.nl

¹ Erasmus School of Law, Rotterdam, Netherlands

² Department of Medicine DIMED, University of Padua, Padua, Italy

³ Erasmus Medical Centre, Rotterdam, Netherlands

⁴ IRCCS San Camillo Hospital, Venice, Italy

⁵ Department of Mathematics, University of Padua, Padua, Italy

⁶ Department of Public, International and Community Law, University of Padua, Padua, Italy

uncertainty whether making data available is subject to the traditional ethical assessment or has to follow a separate research data ethics route. Institutions have different policies and interpretations for that. What is possible at one hospital might be a no-go at another one. Second, researchers will meet with divergent positions that different institutions or ethical committees might have on the assessment of research ethics. Furthermore, using and sharing existing databases in medical research institutions is not unproblematic under the data protection regulation, because the data processing must stay within the purpose for which the data are originally collected; according to art. 5(1)(b) of the General Data Protection Regulation (GDPR)—the primary legal framework for personal data regulation across Europe; in force since 2018.² Clearly, not all databases that could help the research goals are collected for that purpose. Sharing within one or more EU member states may already be risky, because of the divergent policies one might encounter between national Data Protection Authorities. The implementation of the GDPR is a national task, leading to divergent requirements. Crossing the border and sharing with US or Asian data processors is even more problematic because the receiving partners must comply with the GDPR standards and be recognized as such under the extraterritorial effect rules of art. 44 et seq GDPR.³ Finally, in this exemplary casus the question is acute to whom the data belong: Who can claim the sovereignty or ownership over those data? Researchers will have to wait for the hospital management's approval to pool with other databases or open the database for federated learning. The management will add to the requirements already mentioned above a level of risk assessment regarding data- and network security. It is obliged to do so under various legal regimes, not only the provisions against personal data leaks in the GDPR but also the national legislations for vital sectors.⁴

One may wonder where the consent of the patient comes into view. Indeed, according to good clinical practice, the

permission to use the data for care and cure (the primary use of data⁵) will be included in the consent a patient gives when entering into the contractual relationship with the hospital. Leaving aside the concerns one may already have around this consent, evidently the permission did not and could not include the future sharing of data for a research purpose in a time frame and a location yet unknown. Obtaining that permission afterwards is practically unrealistic, where in many cases the treatment relation between patient and hospital will already be terminated years ago, as the patient recovered, moved elsewhere or passed away. Art. 6(4) GDPR requires specific safeguards in this respect, preventing the identification of the patient, also referring to the higher level of protection for health data (art. 9(2)(j) GDPR). This last point significantly obstacles the opportunity to plan retrospective observational studies that often can be drawn for preliminary testing of some hypotheses or for evaluating data from large cohorts of patients (i.e. the big data analysis). The specific measures are repeated in the research exception of art. 89 GDPR, a dedicated provision to enable scientific research with personal data.

That brings us to the GDPR standard for anonymization or pseudonymization, which is perceived as divergent and stricter than those in other jurisdictions of the world [3]. For the MRI brain tumor data in our example, the question arises to what extent one has to anticipate future technological developments, so as to exclude any imaginary type of reconstructing patients' identity from the data. A reasonability test could be expected here but has not been recognized so far across the EU, let alone globally.⁶

The second example reinforces the impression that the GDPR actually obstacles the pursuance of fundamental public interests as we saw above: How to deal with the possible future developments of technology? Releasing an algorithm for use in other medical research environments would not qualify directly as processing personal data. But as soon as the expectation arises that technology may come on the market that can 'read' from the algorithm information on the data it has been trained on, then the question of qualification is again on the table.⁷ Would the data protection level one have to comply with under the GDPR also apply

² 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), (<https://gdpr-info.eu/>).

³ To be joined by art. 52 – 54 of the draft EHDS regulation mentioned above, introducing the authorized participants to HealthData@EU from third countries; only for non-identifiable data (see below).

⁴ E.g. the EU NIS Directive implementation comprises the health sector in Germany, but not (yet) in the Netherlands. Italian legislation has not transposed the NIS Directive into mandatory constraints; it has simply expressed a general obligation to implement an adequate level of protection, embracing a logic of risk management wherein individual actors must autonomously conceive and implement measures that pursue safety objectives. The draft for the revised Directive (NIS2) pursues the harmonization of national regimes, including the health sector in all Member States.

⁵ Defined in art. 2(2)(d) of the draft EHDS regulation as: 'primary use of electronic health data' means the processing of personal electronic health data for the provision of health services to assess, maintain or restore the state of health of the natural person to whom that data relates, including the prescription, dispensation and provision of medicinal products and medical devices, as well as for relevant social security, administrative or reimbursement services.

⁶ The draft EHDS Regulation refers to these measures in art. 44(2) and (3), not referring to actual status quo of technology or a reasonability test.

⁷ As we understand it, the proposed regulation on the European Health Data Space does not address this issue.

now for the shared use of these algorithms? There is no clear guidance there. Such uncertainty can limit the identification of potential biomarkers able to predict the likelihood of disease/response to therapy/prognosis, by using large retrospective data analysis.

The two strands of exemplification in this essay are intended to show how in current research practices the law fails in one important aspect of its purpose: facilitating the activities that are deemed highly desirable in society. As a consequence, there is a kind of unproductive pressure on the climate for innovative data-driven medical research. Therefore, a plea must be made for a fresh view and the development of legal/ethical standards that re-align the interests of the implied stakeholders better than the current antithesis of privacy versus research. It would not only reflect the actual weight attached to the maintenance and further distribution of high-quality health care at affordable cost. It would also reduce the pressure that the current legal approach puts on individual consent. This individual consent is now the security valve, or a catch-all if one wants. However, a better representation of the interests of the individual patients is direly needed. After all, an individual's consent is hardly the lodestar of medical research and of medicine altogether.

Declarations

Conflict of interest The authors declare to have no conflict of interest.

References

1. Wang RS, Maron BA, Loscalzo J (2015) Systems medicine: evolution of systems biology from bench to bedside. *Wiley Interdiscip Rev Syst Biol Med* 7(4):141–161
2. Yaffe MJ (2018) Emergence of “Big Data” and its potential and current limitations in medical imaging. *Sem Nucl Med* 49:94–104
3. European Medicines Agency. “Data anonymisation-a key enabler for clinical data sharing”, London 4 December 2018 *EMA/796532/2018* available on (https://www.ema.europa.eu/en/documents/report/report-data-anonymisation-key-enabler-clinical-data-sharing_en.pdf) accesses 21 Dec 2022

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.