POLICY FORUM

DATA GOVERNANCE

Scrutinizing the EU General Data Protection Regulation

How will new decentralized governance impact research?

By **Luca Marelli**[1*] *and* **Giuseppe Testa**[1,2]

On 25 May 2018, the European Union (EU) regulation 2016/679 on data protection, also known as the General Data Protection Regulation (GDPR), will take effect. The GDPR, which repeals previous European legislation on data protection (Directive 95/46/EC) (*1*), is bound to have major effects on biomedical research and digital health technologies, in Europe and beyond, given the global reach of EU-based research and the prominence of international research networks requiring interoperability of standards. Here we describe ways in which the GDPR will become a critical tool to structure flexible governance for data protection. As a timely forecast for its potential impact, we analyze the implications of the GDPR in an ongoing paradigmatic legal controversy involving the database originally assembled by one of the world's first genomic biobanks, Shardna.

The GDPR set out to harmonize data protection legislation in the EU, with the twofold aim of affording citizens increased protection and empowerment over personal data [Art. 1(2)], while also enhancing the circulation of those data within the EU [Art. 1(3)]. This is geared to provide regulatory support for the establishment of a full-fledged Digital Single Market—a policy cornerstone of the European Commission under President Jean-Claude Juncker. At its core, along with conferring new rights to data subjects [such as the "right to be forgotten" (Art. 17) and the right to data portability (Art. 20)], the GDPR adopts a risk-based, context-specific approach meant to ensure that appropriate data-protection measures are designed and implemented throughout the entirety of the processing activities (as enshrined in the "data protection by design and by default" principle, Art. 25). To this end, the GDPR promotes the responsibility of data controllers [Arts. 5(2) and 24], and it introduces new, decentralized modes of accountability (Art. 40). Additionally, the GDPR lays down specific provisions for the processing of sensitive data (Art. 9) for scientific research purposes (Art. 89), requiring organizational and technical safeguards, such as data pseudonymization, and mandating the designation of a data protection officer in case large-scale and systematic processing of sensitive data occurs (Arts. 37 to 39).

[1]*Department of Experimental Oncology, European Institute of Oncology, 20139 Milan, Italy.* [2]*Department of Oncology and Hemato-Oncology, University of Milan, 20122 Milan, Italy.*
**Present address: Marie Skłodowska-Curie Fellow, Centre for Sociological Research, KU Leuven, 3000 Leuven, Belgium.*
Email: luca.marelli@ieo.it; giuseppe.testa@ieo.it

GOVERNING "BIG-DATA" BIOMEDICINE

Regulatory challenges of big-data biomedicine pivot around (i) the inherently open-ended potential of data, whose digital compatibility makes them valuable for research pursuits that may be wholly disjoined from the original project within which samples or data were gathered, thus undermining the classical rationale for "informed consent"; and (ii) the increasing resolution and scope of data across the full range of digitized human features (from genomes to social networks' logs), with the ensuing and often self-proclaimed erosion of privacy (*2*). Responses have come broadly in two flavors, both aiming for technical fixes, though at different levels of technological engagement.

The first includes attempts to solve the conundrums of the digital age by resorting to yet more complex digitization, as in the recent example of secure multiparty computation that enabled genomic diagnosis while preserving participants' privacy (*3*). The second resorts instead to one of the defining human technologies of our time, that is, governance, meant as the reconfiguration of power structures through procedural architectures and distributed agency (*4*), as in the proposal of trust-building techniques to skirt the zero-sum game of data privacy versus data utility by shifting emphasis from the issue of privacy, per se, to the acquisition of control over data and trust in their holders (*5*).

As one of us has empirically shown (*4*), governance mechanisms have been central to the rise of contemporary biomedicine, as well as to the shaping of European science policy (*6*), by virtue of their mutually reinforcing constitutive tenets: (i) a partial retreat of state powers and governing bodies vis-à-vis the advance of market forces and a plurality of heterogeneous "stakeholders," ushering in a substantial reshaping of decision-making ("decentralization"); and (ii) the increased reliance on soft-rule instruments—such as standards, codes of conduct, and ethical thresholds—in place of more rigid forms of legislative interventions ("standardization"). Both of these features, in turn, have been integral to the structuring of the GDPR.

DECENTRALIZATION

Notwithstanding its binding nature and heavy sanctionatory regime [up to 20 million Euros, or 4% of a company's yearly global revenues, in case of noncompliance (Art. 83)], the GDPR decentralizes by delegating responsibility from national and EU authorities to data controllers (that is, the persons, companies, associations, or other entities that are in control of personal-data processing). As enshrined in the "accountability principle" [Arts. 5(2) and 24], controllers are required to adopt a proactive approach toward data protection and are responsible for the ex ante assessment, the implementation, and the post hoc verification of appropriate measures to ensure and demonstrate that data processing complies with the GDPR.

In providing coarse-grained guidance as to what measures fulfill a controller's obligations, and in making the determination of those measures dependent on the "nature, scope, context and purposes" of the relevant processing (Art. 24), the GDPR is set to promote a controller-based, case-sensitive, and context-specific approach to data protection. This marks a transition from a "paternalistic" to an "autonomy-based" regime in European data protection—in a fashion anecdotally best captured by the remark, overheard at a meeting on privacy law, that "with the GDPR, the EU is living the 1960s of data protection."

The shift toward a decentralized, controller-anchored, and accountability-based model gains salience with respect to secondary research, especially considering the emergence of "dynamic knowledge repositories" proposed as key enablers of "high definition medicine" (*7*). Article 5(1)(b) states that further data processing for scientific research "shall not be considered to be incompatible with the initial purposes" for which personal data were originally collected. In addition, the GDPR introduces criteria for compatibility assessment [Art. 6(4)], to be carried out by the data controller, which aims at ascertaining, on a case-by-case basis,

ILLUSTRATION: SÉBASTIEN THIBAULT

whether the further processing of personal data (without the data subject's consent), is compatible with the initial purpose for which data were originally collected. Factors to be taken into account for this "compatibility test" include "the nature of the personal data, in particular whether special categories of personal data are processed" [Art. 6(4)(c)]; "any link between the purposes for which the personal data have been collected and the purposes of the intended further processing" [Art. 6(4)(a)]; "the reasonable expectations of data subjects on the basis of their relationship with the controller as to their further use" (Recital 50); and "the context in which the personal data have been collected" [Art. 6(4)(b)].

STANDARDIZATION

Such flexibility in data processing extends to the foundational tenet of human biomedical research—informed consent. Although the GDPR requires "specific [and] informed" consent of the data subject [Art. 6(1)(a) and Recital 32], it recognizes that researchers may face the impossibility of fully identifying all potential future research purposes at the time of data collection. Accordingly, Recital 33 states that, if too specific a consent would impinge on the purpose of research, "data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognized ethical standards for scientific research."

This key provision, which intersects still unsettled bioethical debates on the appropriate modes of informed consent, has two major implications. First, dispelling concerns voiced in relation to previous drafts of the GDPR suggesting otherwise (*8*, *9*), it lends the full legislative weight of the GDPR in support of broad consent whenever the criterion of specific consent for specific research use at the moment of data collection proves impossible to satisfy, as in the case of biobanking.

Interestingly, the Article 29 Working Party (the current EU data protection advisory body) recently issued further clarifying guidelines on the notion of consent in the GDPR (*10*). These guidelines reaffirm, as the default option, the requirement for specific consent. At the same time, they avoid a collision course with Recital 33 by treading a thin line between research purposes that, while being required to be "well-described," ought not always be "fully specified." In such cases, additional safeguards that could offset the lack of a specified purpose are recommended (such as provision of a comprehensive research plan before commencement of a project and/or increased transparency on its development to allow participants to exercise their right to withdraw consent). While submitting the flexible approach of Recital 33 to a "stricter interpretation" and "high degree of scrutiny," this interpretative guidance paves the way for controllers to avail themselves of broad consent whenever required by the intended research purposes.

Second, the provision contained in Recital 33 bestows on institutionalized ethics (that is, ethics committees' guidelines, along with other soft-law instruments such as professional codes of conduct, compare also Art. 40) an enhanced role in defining the scope of data processing for scientific research. More broadly, this role is enhanced in establishing general standards of practice that research with human biospecimens has been (as of yet) largely eschewing, given the absence of binding legislative requirements comparable to those regulating clinical research.

This is likely to heighten the relevance of the crucial, though still scholarly undertheorized, policy-making role of review boards and ethics committees. It also places increased emphasis on the ongoing effort lead by BBMRI-ERIC (the EU Biobanking and BioMolecular resources Research Infrastructure–European Research Infrastructure Consortium), involving major research organizations, patient advocacy groups, and industrial representatives, to develop a comprehensive code of conduct for the processing of personal data in health research (http://code-of-conduct-for-health-research.eu/). This code is envisioned as the reference standard in the field, enabling international harmonization in the EU and possibly beyond. However, it remains to be seen whether the concrete implementation of the GDPR will allow such sweeping reach by a single code of conduct. Or whether, on the contrary, the GDPR's enhanced investment in institution-

alized ethics will end up promoting regulatory fragmentation through a proliferation of approaches by local ethics committees.

TESTING THE GDPR

For the data protection–versus–data utility conundrum, the implementation of a flexible regime of governance for European data protection, in place of rigid homogenizing provisions (and notwithstanding the possibility that Member States introduce further provisions for the processing of genetic and health-related data [Art. 9(4)], has ambivalent implications.

On one hand, in addition to controllers' discretionary prerogatives, the GDPR upholds a far-reaching "research exemption" to the strict limitations otherwise imposed on the processing of sensitive data, relaxing requirements for consent [Art. 9(2)(j)], further processing [Art. 5(1)(b)], and data storage [Art. 5(1)(e)] (*11*). The reach of this research exemption is magnified by the adoption, as per Recital 159, of an exceedingly broad definition of activities falling under the rubric of "scientific research," including "technological development and demonstration," "applied research," and "privately funded research." Through the combination of these provisions, controllers such as pharmaceutical, direct-to-consumer genetic testing, and digital technologies companies, claiming to process (sensitive) personal data within the scope of scientific research activities, stand to benefit directly from a major regulatory leeway in favor of data controllers over data subjects (*11*).

On the other hand, the context-sensitive approach entailed by mechanisms such as the compatibility test, as well as the enhanced role assigned to institutionalized ethics, could be seen—and harnessed accordingly—as laying down the conditions for increased protection of data subjects and the promotion of their substantive, rather than merely tokenistic, engagement in research.

This can be exemplified in reference to the landmark first-instance ruling of the Tribunal of Cagliari (Italy) (*12*) that overturned a decision by the Italian Data Protection Authority (DPA) (*13*)—the first time ever by an Italian court—that had halted the activities of the United Kingdom–based Tiziana Life Sciences Plc. with the Shardna SpA database. Shardna was an Italian genomic biobank that stored genetic, health, and genealogical data (the latter collected from municipal and parish records archived over 4 centuries) of around 12,000 genetically interrelated residents from Ogliastra, an isolated region in Sardinia, Italy, known for being one of the world's few "blue zones," areas with a high prevalence of centenarians. Shardna was purchased by Tiziana in 2016 amid arguments in local communities, including research participants seeking to halt the foreign and for-profit acquisition of Shardna's database (*14*).

The decision of the Italian DPA to impose an interim block to Tiziana's processing was made on the basis that Tiziana, as the new data controller, had to (i) inform data subjects "of the change of data controller, and the further data processing for scientific research purposes in the field of medical genetics that the new controller may intend to carry out" and (ii) recollect consent from all data subjects whose information was stored in the Shardna database (*13*). The Tribunal of Cagliari instead ruled such a provision "exorbitant" given that the new data controller "pursues the same purposes of Shardna," namely, "scientific research purposes for which consent had been given" (*12*).

Although both the decision of the DPA and the ruling that overturned it unfolded with Italian data protection legislation still in place (*15*), the appeal judgment is expected to occur under the GDPR regime. Though full disentanglement of its legal complexity is beyond the scope of this paper, we argue that adjudication of this case is poised to revolve around the assessment of whether the change in data controller entails (i) a further processing of personal data, (ii) whose scope is not compatible with the original purpose for which consent had been obtained. Such assessment will arguably lead to the conclusion that it is highly implausible that Tiziana will not conduct further processing on the data set originally assembled by Shardna to advance its own research programs in oncology and immunology, irrespective of whether it (dis)continues the lines of research originally initiated by Shardna.

Assessing the compatibility of such further processing will require unpacking the abstract notion of scientific research (the stated purpose of the processing carried out by both Shardna and Tiziana). This will involve scrutinizing the research endeavors pursued by the two organizations, recognizing the substantial differences not only in their research programs and goals but also in terms of their respective governance arrangements, values, and aspirations. Shardna was established as a locally owned and governed biobank, strongly rooted in Ogliastra's communities (the name itself recalls the Shardana population inhabiting Sardinia in the Bronze Age). As such, it was able to achieve high recruitment rates in the local population (*14*) for its research programs focused on the genetics of multifactorial diseases typical of Ogliastra, and thus of primarily, albeit not exclusively, local relevance (*12*). By contrast, Tiziana is an international, profit-oriented biotech company with a distinct research focus, with feeble ties to the local communities engaged in research. Consequently, irrespective of how "broad" the original consent given to Shardna was, a cogent case could be made that the relationship between data subjects and the controller—a key criterion identified by the GDPR to determine compatibility of further processing—has been considerably altered upon the change in data controller. This could be argued to make further processing incompatible with the original purpose, thus requiring reconsenting of research participants.

Whatever its outcome, this case is an exemplary testing ground because its adjudication is bound to set jurisprudence for two central issues in contemporary biomedicine that the GDPR, in its bottom-up valorization of context, leaves open to interpretive ingenuity: the reuse of personal data upon a change in the property of biobanks and data repositories and what the relevant demarcations to be traced are, in the age of big data, within the category of scientific research in its ever increasing interdependence with other socioeconomic domains. Collective engagement with the versatility of this legislative tool by European scientists and citizens will thus be key to ensure its impact is scientifically and socially robust.

In conclusion, translating into practice the tensions between greater freedom and greater accountability of research defines the very scope of the GDPR and, more broadly, of the EU's experimenting with the governance of technoscience. ■

REFERENCES AND NOTES

1. Regulation (EU) 2016/679 (General Data Protection Regulation) 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, which repeals Directive 95/46/EC; www.eugdpr.org/.
2. Y. Joly *et al.*, *Cell* **167**, 1150 (2016).
3. K. A. Jagadeesh *et al.*, *Science* **357**, 692 (2017).
4. H. Nowotny, G. Testa, *Naked Genes* (MIT Press, 2011).
5. Y. Erlich *et al.*, *PLOS Biol.* **12**, e1001983 (2014).
6. S. Jasanoff, *Designs on Nature* (Princeton Univ. Press, 2005).
7. A. Torkamani *et al.*, *Cell* **170**, 828 (2017).
8. D. Hallinan, M. Friedewald, *Life Sci. Soc. Policy* **11**, 1 (2015).
9. J. Kaye *et al.*, *Eur. J. Hum. Genet.* **23**, 141 (2014).
10. Article 29 Working Party, Guidelines on Consent under Regulation 2016/679, WP259; http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=615239.
11. K. Pormeister, *Int. Data Privacy Law* **7**, 137 (2017).
12. Tribunal of Cagliari, Sentenza n. 1569 (6 June 2017).
13. Italian Data Protection Authority, Garante per la protezione dei dati personali, Provvedimento n. 389 (6 October 2016); www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/5508051.
14. S. Kirchgaessner, "Ethical questions raised in search for Sardinian centenarians' secrets," *The Guardian* (12 August 2016).
15. Decreto legislativo n. 196 (Italy, 30 June 2003).

ACKNOWLEDGMENTS

This work was supported by the EPIGEN Flagship Project of the Italian National Research Council (CNR) (L.M. and G.T.) and the Italian Ministry of Health (Ricerca Corrente to G.T.).

10.1126/science.aar5419

Scrutinizing the EU General Data Protection Regulation

Luca Marelli and Giuseppe Testa

Science **360** (6388), 496-498.
DOI: 10.1126/science.aar5419