Double standards for research? Unintended consequences of Open Access Policies

Deborah Mascalzoni [1] [2] & Heidi Beate Bentzen [3] [4] *, Isabelle Budin-Ljøsne [5], Lee Andrew Bygrave [4], Jessica Bell [6] [7], Edward S. Dove [8], Christian Fuchsberger [1], Kristian Hveem [9] [10], Michaela Mayrhofer [11], Viviana Meraviglia [1], David R. O'Brien [12], Cristian Pattaro [1], Peter P. Pramstaller [1], Vojin Rakić [13], Alessandra Rossini [1], Mahsa Shabani [14] [15], Dan Jerker B. Svantesson [16], Marta Tomasi [17] [18], Lars Øystein Ursin [19], Matthias Wjst [20] [21], Jane Kaye [6] [7]

*Mascalzoni and Bentzen are shared first authors.

[1] Institute for biomedicine, EURAC Research, Bolzano, Italy

[2] Centre for Research Ethics and Bioethics, Uppsala University, Uppsala, Sweden

[3] Centre for Medical Ethics, Faculty of Medicine, University of Oslo, Oslo, Norway

[4] Norwegian Research Center for Computers and Law, Faculty of Law, University of Oslo, Oslo, Norway

[5] Norwegian Institute of Public Health, Oslo, Norway

[6] HeLEX Centre for Health, Law and Emerging Technologies, University of Oxford, Oxford, United Kingdom

[7] Melbourne Law School, University of Melbourne, Melbourne, Australia

[8] School of Law, University of Edinburgh, Edinburgh, United Kingdom

[9] HUNT Research Centre, Department of Public Health, Norwegian University of Science

and Technology, Levanger, Norway

[10] K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway

[11] BBMRI-ERIC, Graz, Austria

[12] Berkman Center for Internet & Society, Harvard University, Boston, USA

[13] Center for the Study of Bioethics, Institute of Social Sciences, European Division of the UNESCO Chair in Bioethics, University of Belgrade, Belgrade, Serbia

[14] Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[15] Leuven Institute for Human Genomics and Society, Leuven, Belgium

[16] Centre for Commercial Law, Faculty of Law, Bond University, Gold Coast, Australia

[17] European Centre for Law, Science and New Technologies, University of Pavia, Pavia, Italy

[18] Faculty of Economics, Free University of Bozen-Bolzano, Bolzano, Italy

[19] Faculty of Medicine, Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway

[20] Helmholtz Center Munich, National Research Centre for Environmental Health, Institute of Lung Biology and Disease, Munich, Germany

[21] Institute of Medical Statistics, Epidemiology and Medical Informatics, Technical University Munich, Munich, Germany

Author Heidi Beate Bentzen's highest degree is Master of Laws (LL.M.), and author David R. O'Brien's highest degree is Juris Doctor (J.D.). All other authors' highest degree is PhD.

Corresponding author:

Professor Jane Kaye

HeLEX Centre for Health, Law and Emerging Technologies

Nuffield Department of Population Health

University of Oxford

Ewert House, Ewert Place, Banbury Road, Summertown, Oxford OX2 7DD

United Kingdom

jane.kaye@law.ox.ac.uk

Telephone: +44 (0) 1865 287898

As part of an agenda of Open Access to data, funding bodies and publishers require researchers to deposit research data, including individual-level sensitive data such as health data, in research repositories (RR).(1) Researchers acknowledge that this is necessary to facilitate and reproduce new discoveries. However, for some researchers, the choice is whether to adhere to Open Access policies or to comply with the law. The deposition of individual-level sensitive data in RRs can contravene regional laws, ethical constrains and undermine the trusted relationship between participants and researchers. It also creates double standards based on national specificities – a two-track system in the research world – jeopardizing collaborations and impacting scientific excellence and evidence-based medicine.

This tension is acute as the European Union's General Data Protection Regulation (GDPR), which started to apply in May 2018, introduced administrative fines of up to EUR 20 million for non-compliance, enforced by active data protection authorities in each EU country.(2) By complying with the law instead of the Open Access requirements, researchers may experience difficulties funding and publishing their work, regardless of its scientific value.

The GDPR's focus on the individual's rights and researchers' transparency and accountability obligations, make it difficult to adhere to current RR policies:

1. Data is frequently considered identifiable, and as such, subject to the GDPR. The Norwegian HUNT population study is an example of the effect of conflicting requirements. In a sub-study, whole genome sequencing of participants was partly financed by the National Institutes of Health (NIH), and genetic data and some phenotypic variables were to be deposited in the database of Genotypes and Phenotypes (dbGaP). In 2011, The Norwegian ethics committee regarded whole genome sequencing data as "genuinely

anonymous", and the risk of re-identification from data in dbGaP as "only theoretical". Hence, the opt-out consent stated that "the genetic information is stored anonymously, i.e. without identifying information."(3) Today, neither the ethics committee nor HUNT consider whole genome sequencing data as anonymous, contradicting the opt-out consent that genetic data is stored genuinely anonymously.

Providing consent to deposit of data in anonymized form is different from consenting to deposit data that are re-identifiable. Genomic data as in the example above, are generally regarded as identifiable in a legal sense, and the deposit of such data will be subject to the GDPR for any EU-based researcher.(4) This calls into question the continued legal validity of past consents for the processing of genomic data. Similarly, one must assess whether other types of data concerning health are identifiable according to the GDPR. If they are, the GDPR explicitly treats data concerning health and genetic data as special categories of personal data, the processing of which must comply with relatively stringent conditions.(5)

2. Personal data must be collected for specified, explicit, and legitimate purposes. Further processing for scientific research purposes is considered compatible with the initial purposes, but the GDPR recognizes an individual's right to object (unless the processing is necessary for reasons of public interest under specific national laws), the right to withdraw consent, and the right to be contacted in case of data breaches.(6)

   This means that information about the uses of data in the RR shall be given to the individual so that they can exercise their rights of objection and withdrawal. Currently RRs rely on Data Access Committees to determine data access requests, which means that institutions may not know how the data is being shared so they cannot advise participants who consequently lose their opportunity to withdraw or object to processing.

3. The collection and retention of data should be adequate, relevant and limited to what is necessary to fulfill the purpose. This is referred to as the data minimization principle in the

GDPR. This does not allow long-time retention for unspecified uses. Currently, RRs do not have mechanisms in place to regularly review the use of the data to establish if it is still needed and to delete the data if it is not.

4. The cross-border nature of RR means that they may be subject to the jurisdiction of several countries' laws, courts and administrative bodies at the same time. Complying with one country's law can mean breaking another country's law. Recent court cases illustrate the legal complexities and the lack of foreseeability third-party access requests entail.(7)

   The GDPR allows EU member states to maintain or introduce their own further conditions on processing of health and genetic data – an opportunity that is exploited by some states. For instance, the deposition of individual genome data is illegal in some countries, introducing a scientific bias based on nationality. The Italian population study SardiNIA/ProgeNIA provides an example of how RR policies may conflict with local law. The study collects extensive phenotypes and detailed genetic data, based on genotyping with high density arrays coupled with low-pass whole genome sequencing, in four geographically clustered villages in Sardinia. The study design and the special features of the population being studied have enabled detection of numerous genetic associations with traits of biomedical relevance.(8) The study is funded by the NIH, and a prerequisite of NIH funding is deposition of genetic data and some phenotypic variables in dbGaP. Thus, about 10 years ago, at a time where there was uncertainty in the scientific community as to the interpretation of the Italian legislation, the project complied with NIH policy to deposit a set of sequencing and phenotypic data in dbGaP. Because the sharing of individual-level data involves a theoretical risk that the identity of the participants and their present and future relatives may be discovered by non-authorized parties, the deposition of such data is now prohibited by Italian law, regardless of deidentification techniques, the purpose of the dissemination, limited access strategies, and individual consent.(9)

Conclusions

Open Access policies and RRs must account for diverse legal constraints and adapt to the legal and ethical landscape to enable EU researchers to share data. Failure to do so will lead to double standards in the access to research resources and engender inequities in the scientific evaluation process, endangering scientific excellence which is a prerequisite for evidence-based medicine.

There are a number of actions which RRs could consider:

- Develop policies ensuring cross-border mutual recognition of equivalent access, including ethical review; reduce system inefficiencies; and respect differences in regulation and governance.
- Use an online dynamic consent approach to enable research participants to make their own decisions regarding the sharing-policy of their data over time, and to modify their decisions if conditions change.
- Adopt tools such as data tagging to indicate authorized uses of the data and analytical tools to run research queries from multiple sites without physically transferring the data.(10)

The GDPR should be seen as an opportunity to review RR policy and develop systems that support the Open Access agenda *and* are legally compliant. Getting this balance right will lead to greater transparency and accountability in the sharing of data and a more efficient and equitable scientific practice.

(1) Taichman DB, Sahni P, Pinborg A, et al. *Data sharing statements for clinical trials – A requirement of the International Committee of Medical Journal Editors*. N Engl J Med 2017; 376:2277-2279. Kiley R, Peatfield T, Hansen J, Reddington F. *Data sharing from clinical trials – A research funder's perspective*. N Engl J Med 2017; 377:1990-1992.

(2) Article 83 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) (hereinafter GDPR).

(3) The HUNT study – a longitudinal population health study in Norway. NTNU HUNT Research Centre. (http://www.ntnu.edu/hunt). Accessed October 1, 2018. HUNTing for myocardial infarction genes using a new whole exome genotyping array. REK Regional Committees for Medical and Health Research Ethics, 2011. Informasjonsskriv og samtykkeerklæringer. NTNU HUNT Research Centre. (https://www.ntnu.no/hunt/deltaker/brev). Accessed October 1, 2018.

(4) Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. *Identifying Personal Genomes by Surname Inference*. Science 18 Jan 2013: 321-324. Hayden EC. *Privacy protections: The genome hacker*. Nature 497, 172-174 (09 May 2013). Article 2 and Recital 26 GDPR.

(5) Recital 34 and Articles 4(13), 4(15), and 9 GDPR.

(6) Articles 5(1)(b), 21(6), 7(3) and 34 GDPR. Furthermore, ethical and in some countries also legal constraints may include the right to be contacted in case of results that trigger the duty of care, see Article 27 Council of Europe Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research, CETS No. 195.

(7) Bentzen HB, Svantesson DJB. *Jurisdictional challenges related to DNA processing in transnational clouds*. In: Svantesson DJB, Kloza D, eds. Transatlantic Data Privacy Relationships as a Challenge for Democracy; European Integration and Democracy Series, Vol. 4. Cambridge, United Kingdom: Intersentia, 2017:241-260. For court case examples, see for instance *Yahoo! v. Belgium*, Belgium Supreme Court decision, Cass. P.13.2082.N and *United States of America v. Microsoft Corporation*, Supreme Court of the United States, 584 U.S. 2018.

(8) See for instance Sidore C, Busonero F, Maschio A, et al. *Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers*. Nature Genetics 2015; 47, 1272-1281. Pala M, Zappala Z, Marongiu M, et al. *Population- and individual-specific regulatory variation in Sardinia*. Nature Genetics 2017; 49, 700-707.

(9) Paragraph 9.9 of the General Authorization No. 8/2014 for the Processing of Genetic Data (Italian Official Gazette, n. 301, 30 December 2014).

(10) Global Alliance for Genomics and Health: Ethics Review Recognition Policy. Global Alliance for Genomics and Health, 13 February 2017. (https://www.ga4gh.org/docs/ga4ghtoolkit/regulatoryandethics/GA4GH-Ethics-Review-Recognition-Policy.pdf). Accessed October 1, 2018. Dove ES, Townend D, Meslin EM, et al. *Research Ethics. Ethics review for international data-intensive research*. Science 2016; 351(6280): 1399-1400. Kaye J,Whitley EA, Lund D, Morrison M, Teare H, Melham K. *Dynamic consent: a patient interface for twenty-first century research networks*. Eur J Hum Genet 2015; 23(2): 141-146. Dyke SO, Philippakis AA, Rambla De Argila J, et al. *Consent Codes: Upholding Standard Data Use Conditions*. PLoS Genet 2016; 12(1): e1005772. Ardeshirdavani A, Souche E, Dehaspe L, Van Houdt J, Vermeesch JR, Moreau Y. *NGS-Logistics: federated analysis of NGS sequence variants across multiple locations*. Genome Med 2014; 6(9): 71. Budin-Ljøsne I, Burton P, Isaeva J, et al. *DataSHIELD: an ethically robust solution to multiple-site individual-level data analysis*. Public Health Genomics 2015; 18(2): 87-96.