Data Sharing Under the General Data Protection Regulation: Time to Harmonize Law and Research Ethics?
Antonia Vlahou, Dara Hallinan, Rolf Apweiler, Angel Argilés, Joachim Beige, Ariela Benigni, Rainer Bischoff, Peter Black, Franziska Boehm, Jocelyn Céraline, et al.

To cite this version:
Antonia Vlahou, Dara Hallinan, Rolf Apweiler, Angel Argilés, Joachim Beige, et al.. Data Sharing Under the General Data Protection Regulation: Time to Harmonize Law and Research Ethics?. Hypertension, American Heart Association, 2021, 74 (4), pp.1029-1035. 10.1161/HYPERTENSION-AHA.120.16340. hal-03147039

HAL Id: hal-03147039
https://hal.univ-lorraine.fr/hal-03147039
Submitted on 19 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Data Sharing Under the General Data Protection Regulation
Time to Harmonize Law and Research Ethics?


ABSTRACT: The General Data Protection Regulation (GDPR) became binding law in the European Union Member States in 2018, as a step toward harmonizing personal data protection legislation in the European Union. The Regulation governs almost all types of personal data processing, hence, also, those pertaining to biomedical research. The purpose of this article is to highlight the main practical issues related to data and biological sample sharing that biomedical researchers face regularly, and to specify how these are addressed in the context of GDPR, after consulting with ethics/legal experts. We identify areas in which clarifications of the GDPR are needed, particularly those related to consent requirements by study participants. Amendments should target the following: (1) restricting exceptions based on national laws and increasing harmonization, (2) confirming the concept of broad consent, and (3) defining a roadmap for secondary use of data. These changes will be achieved by acknowledged learned societies in the field taking the lead in preparing a document giving guidance for the optimal interpretation of the GDPR, which will be finalized following a period of commenting by a broad multistakeholder audience. In parallel, promoting engagement and education of the public in the relevant issues (such as different consent types or residual risk for re-identification), on both local/national and international levels, is considered critical for advancement. We hope that this article will open this broad discussion involving all major stakeholders, toward optimizing the GDPR and allowing a harmonized transnational research approach.

Key Words: biomedical research • data management • Ethics, Research • Government Regulation • informed consent

The General Data Protection Regulation (GDPR) became binding law in all European Union (EU) Member States in May 2018.1,2 As a Regulation, it is, in principle, directly applicable to all EU Member States, superseding existing Member State laws. It, thus, represents a significant step toward harmonizing EU data protection laws.3

The GDPR applies to the processing of personal data across a large range of contexts, including those used in biomedical research and pertaining to vast areas of translational and clinical research activities, although it was probably not drafted with all the latter purposes in mind. As such, it has instigated multiple discussions among researchers and, in several cases, has raised major concerns, particularly, but not exclusively, in the genomics community (reviewed in study by Townend,3 Phillips,4 and Hallinan5). With the present intense need for international scientific data and bio-sample sharing, actively supported by...
major publishers and research funders.\textsuperscript{6,7} GDPR requirements, particularly with respect to specific consent for data and bio-sample sharing, generate confusion and uncertainty and create conflicts between Regulation and Research Ethics.\textsuperscript{8} Under the fear of legal and social sanctions in combination with the threat of huge penalties as a consequence of violating GDPR, scientists have become reluctant to exchange data and bio-samples for secondary research.\textsuperscript{9} Increasingly, data and bio-sample use is restricted (to eg, specific hospitals, research institutions, or regions). In addition, the basic principle of scientific publishing that all relevant research data must be made accessible to ensure transparency and data reuse is frequently violated. Public health emergencies (such as the Ebola and Zika virus outbreaks and especially the ongoing coronavirus disease 2019 [COVID-19] pandemic) exemplify the need to facilitate swift and safe data sharing around the globe, without legal delays.\textsuperscript{10,11}

This article aims to review the main principles of GDPR, in particular, pertinent to consent, as related to biomedical research; in this context, it discusses the practical problems related to data and bio-sample sharing, currently encountered by researchers, perceived to be linked directly or indirectly to GDPR; it also provides suggestions for interpretations and potential adaptations of the Regulation to better fit the practical realities of current biomedical research.

## CONSENT AND PERSONAL DATA IN THE GDPR

Focusing on the most relevant issues, basic definitions and major aspects of the Regulation with direct impact on biomedical research are summarized in Table 1.

Processing of personal data concerning health is a priori prohibited (Article 9). Exemptions can be made based on consent which, as shown (Table 1), involves a clear, affirmative action in GDPR (Article 4.11). This action cannot simply be assumed based on failure to opt-out (eg, by not refusing pre-ticked boxes or default settings). In addition, the scope of consent covers data processing for one or more specific purposes (Article 6). “Specific” is not explicitly defined but interpretations suggest that “the objectives of research, the principal investigator and the project’s duration are specified”.\textsuperscript{12}

Exceptions allowing research without consent exist, based on Member State law for “archiving purposes in the public interest, scientific or historical research purposes” (Articles 6, 9). Alternatively, there is the possibility of obtaining consent for purposes not fully specified in advance (what we would call “broad” consent), related to “certain areas of scientific research, when in keeping with recognized ethical standards” (Recital 33). Nevertheless, there is confusion about the extent for which these exceptions could be applicable.\textsuperscript{13}

Both directly identifiable and pseudonymized data used by researchers should be treated as personal data (Table 1), based on the presumed residual risk of research subject identification, in case of pseudonymization (summarized in study by Rumbold and Pierscionek).\textsuperscript{14} Anonymized data are not affected or regulated by GDPR. To ensure anonymization, simply changing the identity label is insufficient. Multiple techniques are existent for further data perturbation to avoid (re) identification based on the variables listed.\textsuperscript{15}

## CURRENT STATUS IN BIOMEDICAL RESEARCH DATA AND BIO-SAMPLE SHARING

Sharing is an undisputed ethical obligation of any investigator (including those conducting patient-centered research) and is a prerequisite for scientific and clinical advancement.\textsuperscript{16} To frame this basic condition, pertinent ethical norms exist, including study monitoring by independent ethics committees, aiming to protect human subjects, while at the same time, promoting research collaboration (for a historical perspective, see study by Phillips).\textsuperscript{3} The need for sharing in research becomes even more obvious and critical, when studying rare diseases (such as rare cardiac diseases), which requires compilation of resources to allow meaningful research. In addition, and regardless of the disease incidence rates,
the increasing diversity and complexity of molecular profiles and pathophysiologic phenotypes, underscores the need for adopting a systems integrative approach to define associations and causes, and, ultimately, biology-driven biomarkers and therapeutic targets. As an example, meta-analysis of genome-wide association datasets enables the identification of cardiovascular or kidney disease risk factors or confounders, that are too rare to be revealed at the individual study level.20,21 Likewise, fueling artificial intelligence algorithms with sufficiently large, multi-level datasets (clinical, pathophysiologic, imaging, molecular), offers the unique opportunity to understand disease in a holistic manner.20,21 Along the same lines, data linkage can lead to better disease predictors, which was recently shown of particular value in stroke research.22 Such large scale holistic analyses can only be realized via data sharing and data reuse and are actively supported by major medical associations (eg reflected in the establishment of standards for data collection by the American Heart Association/American College of Cardiology Task Force on Data Standards21) and research funders, such as the EU Framework Program Horizon 2020, the National Institutes of Health, United States, and others.

GDPR-triggered restrictions linked to the requirements for specific consent generate uncertainty and slow down or even prohibit research activities. Contributing to this problem are the exponential developments in molecular technologies, which allow high resolution multidimensional analyses of a biological sample, making it hard or even impossible in the case of high resolution genetic information, to truly anonymize study participants.13,23 Having to comply with an extensive (88 pages) legal framework increases the need for legal advice. As a result, collectively, valuable data and clinical samples may remain unused or significantly underused, causing a decrease of the power of the analyses.

To receive guidance on this critical issue of data and bio-sample use and sharing, as linked to study participant consent requirements in the context of the GDPR, several relevant translational networks (eg, the European Uremic Toxin Work Group [EuTox—https://www.uremic-toxins.org/], the European Society of Urologic Research [ESUR; https://uroweb.org/section/esur/information/], the Clinical Biomarkers COST Action [CliniMark; https://clinimark.eu/], and the International Bladder Cancer Network [IBCN; http://ibcnweb.net/]), initiated a discussion on the topic.

**REFLECTIONS ON IMPLEMENTING GDPR IN SCIENTIFIC DATA AND SAMPLE SHARING**

The main questions/problems currently distressing researchers and respective answers deduced from the Regulation and discussions with legal experts are summarized in Table 2.

The answers provided (Table 2) highlight some main impediments perceived to be directly or indirectly imposed by the GDPR on biomedical research data and bio-sample sharing:

(1) Although the initial aim of the GDPR was to harmonize practices in the EU Member States, hence facilitating data sharing, a significant part of decision-making is still left to the Member States, increasing confusion and bureaucratic complexity. As a consequence, implementation of multinational collaborative projects relying on data and bio-sample sharing and mobility frequently face de facto regulatory deadlocks, which in our experience, to the least, slow down progress. Even more restrictions apply in the case of research involving partners outside the EU (including collaborations with European countries, Japan, or the United States), requiring a “lawful basis for making the transfer”.9

(3) Consent to data use ensuring protection of participants, undoubtedly constitutes a pillar of clinical research ethics and the principal condition to obtain ethical approval. However, the stated need for specific consent raises doubts as to the legitimacy of the use of valuable stored data and bio-samples covered by pre-GDPR consent forms. Calling for national regulation to legitimate processing, when the specific consent is not fully available or cannot be readily obtained, opens Pandora’s box in terms of the complexity of the legal landscape. The high complexity in relation to the scope of consent is illustrated by the problems that arise when existing data and bio-samples collected with the purpose of identifying specific biomarkers, for instance, for cardiovascular disease, would be considered as control in a study targeting the identification of specific chronic kidney disease biomarkers. In the absence of an explicit specification of this latter prospect in the original consent referring to cardiovascular disease, the use of such data and samples in chronic kidney disease research may be, a priori, prohibited. A possibility may exist to allow reuse with a supplemental consent. If this is not possible, reuse may be possible if a relevant national law has been enacted. However, in this latter case, use of Member State law to legitimate processing may end up causing issues in cases of collaborations requiring cross-border data and bio-sample exchange involving multiple EU Member States, as national laws do not always align.

(3) A potential solution in the latter case could be the use of data and bio-samples in an anonymized manner, hence, being exempt from GDPR and (re) consent requirements. However, considering the frequent availability of multiparametric, clinical, pathological, and molecular data characterizing a study participant, as mentioned above, it is reasonable to assume that true anonymization is only possible at the cost of significant loss of information (perturbation of data), especially in the context of -omics and
and risks with broad and other emerging forms of consent, namely meta-consent (allowing the research participant to define what type, when and how consent may be provided in the future) or dynamic consent (through web platforms). In addition, it would explain residual risks for reidentification. There is an apparent ever increasing willingness of individuals to share their data for research advancement, reflected in the existence of large personal genome projects (such as the Harvard or Canadian, United Kingdom, Austrian, Korean, and other national personal genome projects; https://www.personalgenomes.org/). Interactions of researchers and patients should be intensified in the form of large patient-linked networks (for example, the European Kidney Health Alliance; involving Kidney Patients and Foundations, Nephrologists and Nurses, http://ekha.eu/; or the US National Patient-Centered Clinical Research Network [PCORNet]), consisting of health care centers, as well as patient organizations (Patient-Powered Research Networks; https://pcornet.org/patient-centric/), and via regular local outreach initiatives led by academic health centers, with the aim of achieving patient-centered informed consent, anticipating as much as possible future developments.

(2) Addressing issues related to secondary use of data and need for respective re-consent is the most challenging problem, as, for the time being, this seems to be in many cases subject to national law. In our opinion, this cannot be solved by an ad hoc administrative, legal, or policy change only; it must be the topic of a broad multidisciplinary discussion involving researchers, patient advocacy groups and the general public, ethics experts, regulators, and politicians targeting clarity and harmonization, as much as possible going beyond national boundaries, in line with the international character of high-level research. An example in this direction forms the design of the Final Rule, the latest update of the federal policy for the protection of human subjects in the United States, which emerged after being subject to commenting by the general public, various agencies, and expert stakeholders. Our proposal is that the lead to open this discussion is taken by a Consortium of several acknowledged societies involved in the field, including patient advocacy groups; these should work closely with relevant authorities at the EU level (EU Directorate General [DG] for Research and Innovation, the European Commissioner for Research, Innovation and Science, and the Committee of Members of European Parliament for Industry, Research and Energy), and eventually also involve the respective national authorities, toward harmonizing interpretations of the law, aiming at minimizing boundaries impeding cross-border research in the EU.

(3) In this context, the generation of a standardized consent form (broad consent allowing generic future research activities), acceptable to all stakeholders involved, would be of enormous benefit. Such efforts

Table 2. Main Practical Questions-Concerns of Researchers and Respective Answers Deduced From the GDPR

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Can data be shared and reused after publication?</td>
<td>Yes, if the prospect of data sharing and further use was included in the original consent form.</td>
</tr>
<tr>
<td></td>
<td>If not specified in the original consent, exceptions may be allowed depending on the inability for re-consenting.</td>
</tr>
<tr>
<td></td>
<td>Obtaining legitimation for exceptions can be complex and will be governed by the GDPR (Articles 9[2][i], [g], and [j] and the national laws of individual Member States).</td>
</tr>
<tr>
<td>(2) Can follow-up (monitoring over time) data be used and shared?</td>
<td>Yes, if the research participant has specifically consented.</td>
</tr>
<tr>
<td></td>
<td>Exceptions may be applicable, as above (point 1).</td>
</tr>
<tr>
<td>(3) Can biobanked data and bio-samples be shared and used?</td>
<td>A positive or negative response by default is not possible. It depends on the specifics of the situation.</td>
</tr>
<tr>
<td></td>
<td>Issue relates to the use of data and samples in broad fields of research, regularly linked to broad consent. GDPR does not use the term broad consent; but is also not entirely opposing to the concept of broad use (Recital 33).</td>
</tr>
<tr>
<td></td>
<td>Major existing international biobanking instruments are supportive of broad consent (eg, the World Medical Association [WMA] Declaration of Taipei).</td>
</tr>
<tr>
<td>(4) Are data and bio-sample restrictions to specific researchers/institutes allowed by the GDPR?</td>
<td>National laws (potentially applicable under the GDPR Articles 9[2][i], or Article 9[4]) frequently differ in their requirements for data and sample use from biobank.</td>
</tr>
<tr>
<td>(5) How can we generate anonymized data for further sharing?</td>
<td>Multiple anonymization techniques exist depending on the type of identifiers and dimension of the dataset.</td>
</tr>
<tr>
<td></td>
<td>Anonymization requires well trained personnel; with awareness of the subject matter and individual importance of each data field.</td>
</tr>
</tbody>
</table>

GDPR indicates General Data Protection Regulation.

Big Data analyses. Furthermore, anonymization may be incompatible with many current research objectives (such as study of disease progression in a longitudinal, individualized manner, as required to establish treatment in a personalized approach).

SOLUTIONS

Specification of the impediments forms the basis for a fruitful discussion toward resolution or improvement. Potential ways forward might include (summarized in Table 3) the following:

(1) Similar to other areas in medicine, public/general population education and engagement in research pipelines is critical. This would target getting a clear understanding of the definitions and associated benefits
in the case of biobanking do exist, for example, in Germany,12 but should be expanded on a harmonized European (and, possibly, international) level. Such a form should inform research participants that complete anonymization may not always be feasible (eg, in the case of whole genome sequencing) and that pseudonymization should be pursued, considering the international dimension of research. This form is intended to serve as a basis for discussion including ethics/legal experts and research participants-patient representatives for further improvement (including abbreviating and simplifying to obtain a legally and ethically sound template, that meets the needs of both research participants and researchers).

(4) The frequently observed restriction of data and bio-samples to specific institutions or countries, although not instructed by GDPR, is often the consequence of fear of violating GDPR. Basic research ethics through implementation of FAIR principles (Findable, Accessible, Interoperable, and Reusable data10) and compliance with privacy and data protection principles should be observed and can be monitored, as recently described.31 Systems, such as DATASHIELD, for simultaneous analysis of multi-source individual data without actual data transfer,32 data hubs allowing highly regulated access,3 or blockchain platforms offering de-centralised data access and interoperability,33 provide data sharing solutions which help maintaining data integrity and patient security. Both the patient and scientific communities would benefit from a harmonized European roadmap on this issue.

**CONCLUSIONS**

The collective experience accrued by now renders the time ripe to assess whether or not the GDPR is on the right track with respect to scientific data and bio-sample sharing and facilitating adherence to the basic research ethics of data sharing. Although, in principle, the GDPR is a step toward regulatory harmonization, based on the issues presented above, adjustments are urgently needed. These should (1) increase harmonization by minimizing differences caused by exceptions based on national laws; (2) seek explicit endorsement of the concept of broad consent and consequently; (3) better define the roadmap for secondary use of data and biosamples at European (and, if possible, international) level. The lead to formulate the working draft should be taken by a team of acknowledged learned societies in the field including patient advocacy groups working closely with experts at the EU level; finalization should be made following a period of commenting by a broad multi-stakeholder audience. This process should evolve in parallel to promoting engagement and education of the public in the relevant definitions (of, eg, broad [specific, meta-, or dynamic] consent; data sharing; residual risk for re-identification), led by academic health centers on a local level and amplified by large patient-centered multidisciplinary networks. We hope that this article will serve as a catalyst for this broad discussion involving all major stakeholders, toward optimizing GDPR to facilitate biomedical research and to produce social benefit and welfare.

**ARTICLE INFORMATION**

**Affiliations**

From the Systems Biology Center, Biomedical Research Foundation, Academy of Athens, Greece (A.V.); FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Eggenstein-Leopoldshafen, Germany (D.H., F.B.); European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom (RA); SAS RD-Néphrologie and Bio-Communication Cardio-Métabolique (BC2M) EA7288 and University Hospital Lapeyronie, University of Montpellier, France (AA); KFH-Nierezentrum und Klinikum St. Georg, Leipzig, Germany (J.B.); Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy (AB, GR); Department of Analytical


