General Data Protection Regulation (GDPR) Research Guidance

August 2021

HARVARD UNIVERSITY
OFFICE OF THE VICE PROVOST FOR RESEARCH

This guidance was originally created by the Johns Hopkins Legal Department. Harvard University has modified the guidance so as to be consistent with University processes, policies and the studies our researchers regularly pursue.

Table of Contents

This guidance assumes some familiarity with the IRB processes, the sponsored research processes (OSP, SPH ORA, HMS ORA, and OTD), the Data Use Agreement (DUA) Policy and the Research Data Security Policy

Please email the relevant office if you have questions related to their review processes

I. General Overview of GDPR

WHAT IS THE GDPR?

In 2018, the European Union (EU) adopted the General Data Protection Regulation (GDPR). GDPR imposes a wide range of obligations, establishes new rights for individuals and is applicable to organizations that control or process personal data of individuals located in the European Economic Area (EEA), regardless of where the organization is located or where the processing takes place.

WHAT COUNTRIES ARE PART OF THE EEA?

Austria	Germany	Malta
Belgium	Greece	Netherlands
Bulgaria	Hungary	Norway
Croatia	Iceland	Poland
Cyprus	Ireland	Portugal
Czech Republic	Italy	Romania
Denmark	Latvia	Slovakia
Estonia	Liechtenstein	Spain
Finland	Lithuania	Sweden
France	Luxembourg	United Kingdom*

GDPR applies to all 27 member countries of the EU, and all countries in the EEA. The EEA is an area larger than the EU, and in addition to the 27 countries in the EU, also includes Iceland, Norway, and Liechtenstein.

*As of January 1, 2021, the United Kingdom (UK) is no longer a member state of the EEA or subject directly to GDPR. The UK has incorporated GDPR requirements into its national laws, and thus this guidance should continue to be used to guide research involving the personal data of individuals located in the UK.

TO WHICH ACTIVITIES DOES THE GDPR APPLY?

GDPR applies to the "processing" of "personal data" by an individual or entity. The term"processing" is extremely broad and generally covers anything that is done to or with personal data, whetherby automated or manual means. This may include collecting (e.g. through surveys, interviews), recording, organizing, storing,altering, consulting, using, disclosing by transmission (e.g. through data use agreements (DUAs), subcontracts, vendor agreements), disseminating or making available (e.g. data repositories, websites, VPN), aligning or combining, restricting, erasing, or destroying data.

Similarly, "personal data" has a broad definition: "any information relating to an identified or identifiable natural person." This includes data that is directly identifiable, such as a name, a unique ID number, physical characteristics, location information, or an IP, email or physical address. The definition also uses the term "identifiable," meaning personal data encompasses information capable of indirectly identifying an individual. This includes pseudonymized personal data (i.e. coded data), and data that could be re-identified by a third party (e.g. a collaborator, data provider, vendor).

CAN THE GDPR BE APPLIED TO ORGANIZATIONS LOCATED OUTSIDE THE EEA/UK?

Yes. GDPR applies to any organization that is established within the EEA and processes personal information in the context of the activities of its establishment in the EEA, even if the processing activity takes place outside the EEA. GDPR also applies extraterritorially to entities (organizations or individuals) that are not established in the EEA, that process the personal data of an individual who is physically located in the EEA, if the entity either (i) offers goods or services to such individual, regardless of whether or not such individuals are charged a fee for such goods or services, or (ii) monitors the behavior of such individual insofar as the behavior takes place within the EEA. These two categories may be interpreted to apply to certain research activities, for example, providing an investigational drug or device to data subjects, monitoring patients longitudinally through repetitive transfers of personal data from the EEA, or tracking the location of a participant via an app or GPS. GDPR is agnostic as to the citizenship of an individual, but rather applies based on their physical presence in an EEA country or the presence of the entity processing the personal data in an EEA country.

There are two different types of data-handlers to which the regulation applies: "controllers" and "processors."A controller is an entity that "determines the purposes and means of processing of personal data," often a principal investigator or primary research site, someone substantively contributing to the research. A processor is an entity that "processes personal data on behalf of the controller," for example a survey or software company, or another type of vendor or contractor that does not exercise discretion over the scope of work. The difference is comparable to the distinction between a subrecipient and a contractor under a sponsored award.

WHAT IF THE DATA IS DE-IDENTIFIED?

Unlike HIPAA, the Common Rule, and other U.S. research and privacy laws, GDPR does not outline specific methods to "de-identify" data by reference to removal of a specific list of identifiers. Rather, the regulation provides that data may be "anonymized" or "pseudonymized."

Anonymization of personal data refers to a de-identification process whereby direct and indirect personal identifiers have been removed and technical safeguards have been implemented such that data cannot be

re-identified. GDPR does not apply to anonymized data, as anonymized data cannot be linked to an identified or identifiable person. Anonymization is judged on a facts and circumstances basis taking into account all the means reasonably likely to be used, such as singling out (i.e. the possibility to isolate some or all records which identify an individual within a dataset), either by the controller or by another person, to identify the person directly or indirectly. Given this definition, anonymization is an extremely high standard that is difficult to meet in practice. A dataset that is de-identified under HIPAA or no longer considered "identifiable private information" for purposes of the Common Rule may not necessarily be anonymized for purposes of GDPR.

GDPR defines pseudonymization as "the processing of personal data in such a way that the data canno longer be attributed to a specific data subject without the use of additional information, provided thatsuch additional information is kept separately and is subject to technical and organizational measures toensure that the personal data are not attributed to an identified or identifiable natural person." Therefore, under GDPR, pseudonymized data refers to data from which identifiers in a set of information are replaced with artificial identifiers, or pseudonyms (i.e. coded data), that are held separately and subject to technical safeguards. Pseudonymized data must be treated as identifiable personal data, and GDPR does apply to the processing of such data. Personal data (even if pseudonymized) is at least Security Level 3.

Some common examples of pseudonymized data at Harvard are:

- *Dr. A is collaborating with a French hospital, and has requested diagnostic information, without direct patient identifiers, from the hospital's ALS patients' medical records*. Dr. A will never know the identity of the patients, however the hospital would be able to identify patients by comparing the provided diagnostic information to the corresponding medical record.
- *A PI receives information under a Data Use Agreement (DUA) on voting records and basic demographics from counties in southern Ireland*. Neither the data provider nor the PI has access to names, however based on certain data points the PI received (age, occupation, county), in several smaller counties, it would be possible to link each data subject to their voting record and identity by referencing additional publicly-available information on the county clerk's website.
- *A professor hires a survey company to disseminate his survey to the company's clientele in multiple European countries. The company will sort the responses using the participant ID number and send all the resulting information to the professor*. Although the professor will not have access to participants' contact information or identifiers, the ID number would allow for the survey company to reidentify the participant, and therefore functions as a link.

HOW CAN PERSONAL DATA BE USED?	
Organizations that process personal data are required to demonstrate that they have a "lawful basis" under GDPR to do so. GDPR specifies acceptable lawful bases for processing of personal data: these include (among others) processing with the informed consent of the data subject and processing for "historic or scientific research purposes." In addition to establishing a lawful basis for processing, researchers may not process certain "special categories" of personal data unless they can demonstrate that one of the exceptions set forth in GDPR applies to permit such processing. Express consent of the data subject is an exception that allows for the processing of special category personal data (though this exception may not always be feasible for processing special category data for secondary research).	
Consent	Personal data can be used for research with the freely given, specific, informed, unambiguous, express consent of the individual data subject. Consent documentation must include a "well-described purpose" for the research and must be clearly distinguishable from other matters (for example, if the researcher would like to retain contact information or responses for a future study, this must be explicitly called out and consented to). The consent also must include a description of all parties who will have access to the personal data (collaborators, vendors, etc.), any special categories of personal data, as well as an anticipated timeline for data destruction or anonymization. (More context below, in "privacy notice" description.) *Note:* Absent a specific exigency, all personal data should be processed and stored using Harvard (or a vetted co-controller's) resources. This enables the University to readily respond to data subjects' questions and provide them with information or access to their personal data, as appropriate.
Third Parties	A controller or processor located in the EEA who wishes to transfer personal data to a jurisdiction outside the EEA that the European Commission has deemed to not adequately protect personal data – which includes transfers to the U.S. – must establish a lawful basis for making the transfer. This is a separate requirement from the requirement to establish a lawful basis for processing data discussed above. The requirement applies to the transfer of personal data to Harvard by an entity in the EEA under DUAs, collaboration agreements, subawards, contractor agreements, or any other relationship that involves the transfer of personal data. Among GDPR's designated 'appropriate safeguards' that satisfy this requirement, the University most often receives personal data in accordance with the standard data protection clauses adopted by the European Commission, the "Standard Contractual Clauses" (the "SCCs"). The SCCs require that personal data be anonymized or destroyed at the end of the project's period of performance (if not before).

	Researchers should reflect this requirement in their anticipated timeline and documentation (e.g. research protocols, data security plan, DUA). *Note:* At Harvard, the SCCs can only be signed by certain institutional signatories, not researchers. Please refer to the Policy on Negotiating and Signing Authority to determine the appropriate signatory. If you have questions about the type of overarching agreement that may be necessary, please utilize the Web-Based Agreement Identification (WAIT) Tool.
Special Categories of Personal Data	Special categories of personal data may require additional data security measures, and must be explicitly described in both internal documentation, and external documentation provided to data subjects (such as consent forms). It is especially important that such data is processed using University resources whenever possible, given the heightened accountability. The special categories are comprised of personal data pertaining to: racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic or biometric information, health information (mental or physical, regardless of source), and sex life or sexual orientation. Personal data relating to criminal activities are similarly restricted.

WHAT INFORMATION SHOULD BE PROVIDED TO THE SUBJECT?

A controller must provide the data subjects with comprehensive information relating to the overarching project. This information is referred to as the "privacy notice" and must be: (i) concise, transparent, intelligible, and easily accessible; (ii) written in clear and plain language; and (iii) free of charge. Generally, the privacy notice must answer the who/what/why/where/when/how questions related to data collection and use:

- What information is being collected/processed?
- Who is collecting/processing it (including contact information)?
- How will it be collected and processed/used?
- Why is it being collected/processed, including the lawful basis?
- How will it be stored and for how long?
- Who will it be shared with (including third parties)?
- What will be the effect of this on the individuals concerned?
- Is the intended use likely to cause individuals to object or complain?
- Will it be transferred from the EEA to a third country (such as the U.S.) and, if so, what is the lawful basis for such transfer?
- The data subjects must also be informed of their rights to request access, rectification, erasure or restriction of processing, to object to processing, and the right to data portability.

In the context of consented research, such notice is often built into the consent documentation. The ESTR Library contains templates for the GDPR Consent Addendum (attached to the 'Adult Consent Form Template'), which is appropriate for current studies, as well as the GDPR Notification, which is appropriate for use with certain studies that pre-date GDPR. Please consult with your IRB prior to sharing documentation with study subjects.

II. SAMPLE SCENARIOS AND APPLICATION OF GDPR

For more information on the internal reviews described below, please reference the HRDSP Applications Summary with Order of Reviews

SCENARIO 1: Harvard is contributing research data to a university in Spain	
Facts	A Spanish university (SU) is conducting a study on juvenile diabetes, and has asked Dr. B for data from one of her ongoing diabetes projects. Dr. B will send U.S. participant data to SU. She will not access any personal data, and will not contribute substantively to the work at SU beyond providing SU researchers with de-identified data on American patients.
Analysis	**Dr. B is not directly subject to GDPR** Dr. B is not providing goods or services to, or monitoring the behavior of subjects in Spain. Additionally, she is not processing personal data on behalf of SU, so is not a processor. *Note*: SU is subject to GDPR because it has an establishment in the EEA and is processing the data provided by Dr. B in the context of such establishment. It does not matter for this analysis that the data provided by Dr. B all pertain to U.S. residents. So, SU may ask Dr. B to revise her consent forms and/or provide a notice to subjects to comply with GDPR so that SU's processing of that data in Spain is compliant with GDPR.
Agreements and Information Security	This scenario would require Dr. B to submit an outgoing DUA request in the Agreements Application, to cover the data transfer to SU. She also must request a corresponding Information Security review in the Data Safety Application.

SCENARIO 2: Professor C has received a grant and is collaborating with German research institutions	
Facts	Prof. C received a grant to research effective remote teaching methods used in the US and Europe and will issue subcontracts to two German research institutions to collaborate on the work. Specifically, the German researchers will collect personal data from professors and students located in Germany and Italy, and send the coded data to Prof. C for analysis and eventual publication in both US and European trade journals. The German institutions will also send reports back to the schools that participated in the study.
Analysis	**GDPR applies to Prof. C as a controller** The German research institutions are subject to GDPR as establishments in the EU who are collecting personal data from subjects located in Germany and Italy, and must

	provide appropriate notice to participants. Additionally, by receiving, analyzing and publishing coded (i.e. pseudonymized) data that could directly benefit the participants' careers and education, it's possible – depending on the specifics of the in-country engagement – that Prof. C is providing research-related services to the European data subjects. *Note*: Prof. C is a controller, and subject to GDPR, even though he will never have access to identifiers. This is both because: (1) the data is pseudonymized, as it could be reidentified by the German collaborators, and (2) as a controller, Prof. C is responsible for processing activities that are being done as part of the project he is leading.
Agreements and Information Security	The German collaborators will need a basis under GDPR to transfer data to Prof. C. in the U.S. This may involve the inclusion of the SCCs in either the subcontracts to the German institutions, or in separate agreements (such as DUAs) with the German institutions. Prof. C is responsible for ensuring that his local administrators and sponsored research office are aware that the project involves data being exchanged with entities in Europe. He also must request a corresponding Information Security review in the Data Safety Application.

SCENARIO 3: Harvard is a study site for a sponsored research project involving human samples collected by EU collaborators	
Facts	A sponsored study's prime recipient has asked Dr. D to perform skin biopsy reads on human tissue samples collected by one of the study's EU sites. The prime recipient sent the HIPAA de-identified biopsy samples to Dr. D for analysis and feedback. At the conclusion of the project, the prime recipient and all the study sites will co-author a paper, and Dr. D will keep the samples for her own secondary research purposes.
Analysis	**GDPR will apply to both the original use and analysis of the data, as well as the data Dr. D will use for her secondary research purposes** Harvard is a controller, because Dr. D is processing and analyzing tissue samples as part of her research that can be reidentified by the EU site and potentially by the prime recipient. She is also contributing substantively to the research, and will co-author a resulting publication. For any secondary uses, Harvard will also be a controller and will need to obtain (or confirm that the EU study site has obtained) express consent from the study subjects for the planned uses. *Note:* Under GDPR, certain data generated from biospecimens cannot be anonymized and remain subject to GDPR even if de-identified per HIPAA standards (for example, if analysis of the specimens would produce results that could be linked to an individual).
Agreements and Information Security	This scenario would require that the SCCs either be included in the subcontract from the prime recipient, or in a separate agreement (such as an MTA) with the prime recipient. Dr. D is responsible for ensuring that her local administrators and sponsored research office are aware that the project involves an exchange of personal data and materials. She also must submit a request for a corresponding Information Security

	review in the Data Safety Application.

SCENARIO 4: Professor E engages in a long-term study that requires continued monitoring of participants when they return home to Europe	
Facts	Prof. E is conducting a study in which student participants will be physically present at Harvard during the initial steps of the study, at which point they will learn how to use the associated phone app. Upon graduation, the participants will return home (potentially in Europe), and Prof. E will continue to monitor certain data points via the app, including location and daily behaviors such as time spent at work, workouts and sleep schedule.
Analysis	**GDPR will apply to the study as Prof. E is monitoring the behavior of EEA residents that takes place in the EEA** Prof. E is a controller under GDPR because he is responsible for the study and the resulting personal data.
Agreements and Information Security	Participants in the EEA will need to be provided appropriate notice (likely as part of the consent documentation). Prof. E must request Information Security review in the Data Safety Application. As there are no third parties involved, no research agreement is required. (If the app is maintained or owned by a third party, and they will have access to the data, Prof. E would need to work with Strategic Procurement and HUIT to review this component, including the underlying agreement.)

III. FLOWCHART: DOES GDPR APPLY TO MY RESEARCH PROJECT?