Guidance: General Data Protection Regulation (GDPR) and Research

Contents

The purpose of this guidance is to help researchers with a basic understanding of how the General Data Protection Regulation (GDPR) may affect their research. It does not include an in-depth analysis of the requirements of the GDPR. For robust training on GDPR, please refer to CITI course ***GDPR for Research and Higher Ed*** available in Yale's portfolio.

1. Overview of GDPR

The General Data Protection Regulation (GDPR) is a European law that went into effect on May 25, 2018. It establishes protections for privacy and security of **personal data** about living individuals in the European Economic Area (EEA). That includes countries in the European Union and certain countries outside of the EU who agreed to implement GDPR. Personal data, including data collected as part of a research study, collected in or transferred from any of the EEA countries is subject to the GDPR.

The purpose of the GDPR is to provide individuals with more control about how their data are collected, used, and shared. It does so by giving individuals in the EEA rights such as rights to access, amendment (rectification) and erasure (**the right to be forgotten**). Organizations that collect or use **personal data** must have lawful justification for why they collect, use, disclose, destroy, or process **personal data**, including data collected for research. The law assigns certain responsibilities to the **controllers** and **processors** to ensure data security using encryption and other technical safeguards. Similar to data protection laws in US such as HIPAA, GDPR requires notification to data protection authorities and affected individuals following the discovery of a **personal data breach**.

2. Definitions

Personal data – refers to any information that relates to an identified or identifiable natural person (i.e., an individual, not a company or other legal entity), otherwise known as a "data subject." Examples of "personal

data" include: a person's name, email address, government-issued identification, or other unique identifier such as an IP address or cookie number, and personal characteristics, including photographs.

Sensitive Data – category of personal data which merit a higher level of protection due to their sensitive nature and risk for greater privacy harm. This includes: information about a data subject's health, genetics, race or ethnic origin, biometrics for identification purposes, sex life or sexual orientation, political opinions, religious or philosophical beliefs, or trade union membership.

Controller – the individual or organization that determines the purpose and means of processing personal data under the GDPR. This includes deciding why data is collected, what data is collected, how it is collected, and who has access to it.

Processor – an entity or person that processes personal data on behalf of a data controller. Unlike a controller, a processor doesn't decide the "why" and "how" of the processing but follows the instructions of the controller. Key obligations for a processor include having a legally binding contract with the controller, implementing technical and organizational measures for data security, and assisting the controller with audits and data subject requests.

Data Subject – an individual whose personal data is collected, processed, or otherwise used by a data controller or processor. This data can be directly or indirectly identifiable through various means, including names, identification numbers, location data, or other specific identifiers.

Personal data breach – breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed.

Pseudonymization – involves processing data so it can no longer be attributed to a specific data subject without additional information.

2 Research activities subject to the GDPR

Research data collected by Yale investigators are subject to GDPR in any of the following circumstances:

- Personal data is collected from one or more research participants physically located in the EEA at the time of data collection (regardless of residency status or nationality of the participant),
- Personal data is transferred from an EEA country to a non-EEA country (like the U.S.) for research purposes,
- Personal data is collected in the US for research on behalf of the **controller** (e.g., a study sponsor) based in the EEA country where the personal data are then transferred to that country.

Personal data from EU citizens who are physically located within the U.S. at the time of data collection are not subject to the GDPR (unless the data are collected on behalf of the controller based in EEA). Data collected in the EU that does not meet the definition of personal data is not subject to the GDPR.

Online research surveys will be subject to the GDPR if they specifically target the enrollment of EEA residents.

3. What Data Is Considered Identifiable Under GDPR

Personal data relates to an identified or identifiable person, including indirectly identified data such as coded data. When data is considered identifiable differs from US regulations for human subjects research and other data protection regulations that are limited in scope to specific areas such as HIPAA or FERPA.

The GDPR definition is more stringent. Data may be considered personal even if it does not rise to the level of what would be identifiable in the US.

Law Regulation	What data is considered identifiable	When data is no longer consider identifiable
Common Rule	Identifiable information is information for which the identity of the individual is or may be readily ascertained by the investigator or associated with the information.	Data is not considered identifiable when the identity of the individual cannot be readily ascertained by an investigator conducting the study.
HIPAA	Identifiable information is information that identifies an individual by including one of the 18 HIPAA identifiers or with respect to which there is a reasonable basis to believe the information can be used to identify an individual.	Data is considered de-identified when i) none of the 18 defined HIPAA identifiers are included in the data (Safe Harbor); ii) data was deemed deidentified via a statistical analysis (expert determination).
GDPR	Data that relates to an identified or identifiable natural person. An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. That likely includes pseudonymized data (e.g., coded data), even where an investigator does not have access to the key-code required to link data to an individual data subject.[1]	Data that has been anonymized i.e., where there is no key-code in existence anywhere to re-identify the data. In other words, the anonymization must be irreversible for the data to no longer be considered identifiable.

3 Lawful Basis

To collect, use, or process personal data subject to GDPR, an organization/individual must have a documented lawful basis.

Lawful Basis	Description	Applicability to Research/Examples
A valid contract with the individual	For example, an employment contract or a contract to supply goods or services.	Unlikely to relate to research.
A legal obligation	The organization could be legally required to process the data.	Unlikely to relate to research.
Vital interests of the data subject	Processing is necessary in order to protect the vital interests of the data subject or of another natural person.	A healthcare research team collects personal health data during an emergency response to protect the life

[1] For more information about pseudonymization, see EDPB Guidelines 01/2025 on Pseudonymisation.

or of another natural person		or physical well-being of individuals in a crisis situation.
Task in a public interest (Public Task)	This includes official functions or tasks in the public interest. For example, schools and other educational institutions, public authorities such as government departments, hospitals, and law enforcement agencies.	Health related research such as public health, archiving purposes in the public interest including scientific or historical research purposes or statistical purposes, humanitarian purposes, including for monitoring epidemics and their spread or in situations of humanitarian emergencies, in particular in situations of natural and man-made disasters.
Legitimate interests	An organization can process data if it has a valid reason, such as fraud prevention or improving services, that is not overridden by the individual's fundamental rights and freedoms. To determine if an interest is legitimate, a three-part test is used: identify a legitimate interest, show the processing is necessary, and balance the organization's interests against the individual's rights and freedoms.	A public health study to investigate the effects of a new treatment for a disease if: • The research objective is significant and in the public interest. • Personal data is necessary for the reliability and validity of the research findings. • There are appropriate measures in place to protect patient data and minimize privacy impacts. • Patients are informed about the data collection and processing, including their rights.
Consent	If the data subject agrees to the processing of their data, after being given when a clear and honest explanation of the reason for its collection and what it will be used for.	A university department conducting behavioral research obtains explicit, informed consent from participants to study their habits and preferences.

4 Sensitive data

Special categories of personal data (sensitive data) include data that reveals "*racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural's sex life or sexual orientation*"[2]. In the context of research, sensitive data can be collected:

- with explicit GDPR compliant consent,
- when processing is necessary for reasons of public interest in the area of public health, or
- when processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.

[2] Art. 9 GDPR: Processing of special categories of personal data

5 Steps to ensure that research complies with the GDPR

- Determine whether your research data is subject to GDPR:

  - Are you processing personal data collected from one or more research participants physically located in the EEA at the time of data collection (regardless of residency status of the participant)?
  - Are you processing personal data transferred from an EEA country to a non-EEA country (like the U.S.) for research purposes?
  - Are you processing personal data collected in the US for research on behalf of the controller (e.g., a study sponsor) based in the EEA country where the personal data are then transferred to that country?
  - Are you processing personal data outside of EEA but targeting enrollment of EEA data subjects?

IF YES to any of the above:

- Verify whether the personal data you are collecting/processing is considered **sensitive data**:

Racial or ethnic origin	Religious or philosophical beliefs	Processing of genetic data	Biometric data for the purposes of unique identification
Political opinions	Trade union membership	Health data	Sex life or sexual orientation information

- Identify the lawful basis for processing the data:
  - Start with ***legitimate interests*** and ***public interests*** AND if you are collecting/processing sensitive data, ***public interest in the area of public health*** or ***archiving purposes in the public interest, scientific or historical research purposes***.

  - **Consent may not be the most appropriate lawful basis for research. It should only be used when no other lawful basis can be applied.** [3]

- Collect only the absolute minimum personal/demographic data needed to complete the study. If your study can be completed using only de-identified data, then we strongly advise you to take this approach.

- Many online survey sites collect personal information, including IP addresses, by default. Ensure that you set up your study to receive **only** the information you are seeking. To the extent possible, verify that any third-party website or app being used for data collection is GDPR-compliant.

- Unless impossible or impracticable (involving disproportionate effort), be transparent with participants about the data collected about them and what the data will be used for. Most likely, your research already includes these measures.

[3] See EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research

6 Consent Used as Lawful Basis

It is recommended that researchers rely on a different primary lawful basis than consent. In rare instances, where consent is the only lawful basis available, the consent form must meet the following requirements:

1. Consent records, including time and date of consent, must be maintained for each subject. In the case of verbal, online, or any other type of undocumented consent, the Principal Investigator is responsible for maintaining a consent log indicating each subject (either by name or study ID number) and the date and time that consent was provided.
2. Consent must be explicit. If the consent form or consent script serves multiple purposes (e.g., a consent form that is also the recruitment email), then the request for consent must be clearly distinguishable.
3. Each subject has a right to withdraw consent at any time. Each subject must be informed of this right prior to giving consent. Withdrawal of consent must be as easy as giving consent.
4. Consent must be an affirmative action. This means that opt-out procedures are not permitted. Use an active ("opt-in") informed consent. Following an informed consent description, a "Click next to proceed to the survey" button or equivalent is sufficient for "active" consent for online data collection.
5. Consent information must be provided in clear and plain language in an intelligible and easily accessible format.
6. For activities in which identifiable data is collected, you must have an executable plan to remove data in the event a participant requests to have his/her data removed. Otherwise, participants must be informed that it is not possible.
7. Consent must be freely-given. Individuals in a position of authority cannot obtain consent, nor can consent be coerced. This means that faculty members or teachers cannot obtain consent from their own students.
8. Consent forms must contain the following information:
    - The identity of the Principal Investigator;
    - The purpose of data collection;
    - The types of data collected, including listing of special categories: Racial or ethnic origin; Political opinions; Religious or philosophical beliefs; Trade union membership; Processing of genetic data; Biometric data for the purposes of unique identification; Health data; and/or Sex life or sexual orientation information;
    - The right to withdraw from the research and the mechanism for withdrawal;
    - Who will have access to the data;
    - Information regarding automated processing of data for decision-making about the individual, including profiling;
    - Information regarding data security, including storage and transfer of data;
    - How long data will be stored (this can be indefinite);
    - Whether and under what conditions data may be used for future research, either related or unrelated to the purpose of the current study.

7 Additional GDPR Requirements

Data Incidents: GDPR has a much shorter time window to report a data breach to the relevant authorities than HIPAA. In the event of a data incident that may qualify as a reportable data breach as defined in the

rule, notify the Human Research Protection Program immediately (HRPP@yale.edu) so that appropriate steps can be taken by the University. Please note that the HRPP in consultation with the Privacy Office will make the official determination as to whether an event qualifies as a reportable data breach. Investigators should not report incidents to EEA data authorities but should work with the HRPP to meet reporting requirements.

Data Subjects Rights Requests: Research participants have a number of rights in the data we may collect in research, including rights to access, amend, and erase their data. The HRPP and/or Privacy Office can assist and provide guidance on responding to rights requests from participants.

Further dissemination of data subject to GDPR: Sharing data subject to GDPR with additional entities such as collaborators and venders may require additional contracting language be included to ensure that the data recipients meet their obligations as a subprocessor under GDPR. The Office of Sponsored Projects and/or the Privacy Office can assist in determining whether additional contractual language is required when further sharing personal information.

8 Revision History

Date	Description
11/10/2025	Initial Effective

Appendix 1: List of European Countries and status of the GDPR

European countries that adopted GDPR:

- Austria
- Belgium
- Bulgaria
- Croatia
- Cyprus
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Ireland
- Italy
- Latvia
- Lithuania
- Luxembourg
- Malta
- The Netherlands
- Poland
- Portugal
- Romania
- Slovakia
- Slovenia
- Spain
- Sweden
- United Kingdom (UK adopted a very similar data protection laws to GDPR, referred to UK GDPR)

Countries in Europe that have not implemented the GDPR regulation:

- Albania
- Belarus
- Bosnia and Herzegovina
- Kosovo
- Moldovia
- Montenegro
- North Macedonia
- Russia
- Serbia
- Switzerland (Although Switzerland does not follow GDPR, it adopted a similar data protection law - Personal Data Protection Law)
- Turkey
- Ukraine