

General Recommendations on the Storage and Use of Data Collected from Research Involving Vulnerable Groups

1. Introduction

Research involving vulnerable groups requires familiarity with legal and ethical issues as well as heightened ethical sensitivity, as participants may experience unequal power relations, limited autonomy, increased exposure to stigmatization, or risks that extend beyond the research context itself (UKRI, 2025; European Commission, 2021). In such circumstances, ethical research practice is based on submitting an application to the appropriate ethics committee before commencing fieldwork, providing clear and accessible information to research participants, ensuring voluntary participation, and implementing a consent process that clearly explains the limits of confidentiality and the anticipated future uses of the data (UKRI, 2025; European Commission, 2021).

At the same time, research teams should comply with rigorous data protection standards and implement the principle of data protection from the design stage onward. This principle should be applied at all stages of the project life cycle, including: planning lawful data processing, data minimization, ensuring secure data processing, and the timely deletion or archiving of data (European Commission, 2021; European University Institute, 2022). Since data from participants belonging to vulnerable groups are more likely to constitute sensitive data or may become such (special categories of data), researchers should apply enhanced safeguards and ensure that the justification for collecting such data is necessary and proportionate (European Commission, 2021; ICO, 2024).

In cases where data processing may involve a high risk to the rights and freedoms of individuals—which is common in research involving vulnerable groups—research projects should anticipate the need to conduct a structured risk assessment and implement appropriate safeguards, including those related to the processing of data for scientific research purposes and responsible data governance (European Commission, 2021; European Data Protection Board, 2021).

Finally, transparency also concerns the ways in which research findings and datasets are shared. Access to data should be “as open as possible,” while restrictions should be justified and proportionate where necessary to protect individuals and communities (UNESCO, 2021; ALLEA, 2023). Where data sharing is appropriate, careful data governance—such as secure processing environments, controlled access, and

advanced anonymization or pseudonymization methods—helps preserve the scientific value of the data while protecting research participants (European University Institute, 2022; ENISA, 2021).

2. Key Definitions

Vulnerable groups: Individuals who may have a limited ability to provide fully informed consent or who may be more exposed to harm or exploitation due to age, disability, health status, socio-economic situation, migration status, institutionalization, or other contextual factors.

Personal data: Any information enabling the direct or indirect identification of a living natural person.

Special category data: Sensitive data requiring additional protection under the GDPR.

Pseudonymization: The processing of data in such a way that it cannot be attributed to a specific individual without the use of additional information stored separately.

Anonymization: The irreversible removal of identifying information, as a result of which the identification of individuals is no longer possible.

3. Some Preconditions

- Personal data are collected only after obtaining the full consent of research participants to participate in the study. Participants must be aware that they may withdraw from participation at any time and that participation is entirely voluntary.
- Participants must receive full information about the purpose of data collection (e.g., in the form of information materials distributed to participants), in a language and format appropriate to the capacities of potential participants, and within a time frame that allows them to read and understand the information.
- General or imprecise wording should be avoided. Instead, researchers should specify as clearly as possible the purpose of data collection and the scope of their accessibility to third parties.
- Ethical consent and consent under the GDPR – the former concerns participation in the research, while the latter refers to consent for the processing of personal data. Each of these should be clearly distinguished and explicitly defined (Wyczik, 2025).

- Consent under the GDPR – in this regard, researchers should explain how the collected data will be processed, who will have access to them, and for what purposes they will be used (Wyczik, 2025).

4. Data Security, Storage, and Use

Anonymization

It is a process aimed at protecting the identity of research participants by making it inaccessible and impossible to determine. After effective anonymization, the data must still retain scientific value (i.e., provide information relevant to the given research field) (Wieczorek, 2025). This constitutes a mandatory stage following the completion of data collection and prior to any use or public dissemination of the data.

To this end, in addition to removing all names and surnames, researchers should:

- seek to minimize as much as possible the scope of personal data disclosed about research participants;
- exercise particular caution regarding any information that could enable the identification of participants, which represents a particular challenge in research involving vulnerable groups.

If anonymization requires the removal of substantial portions of the text, resulting in the text becoming unintelligible, this constitutes a significant reduction in data quality and may even amount to the destruction of the data. If anonymization renders the data unusable, it should not be carried out. In such cases, the data should not be publicly shared, and only metadata should be made available.

Data Management Plan (DMP)

A Data Management Plan is a document specifying details concerning the collection, processing, protection, storage, use, and reuse of research data. This document is dynamic (“living”) and should be updated (versioned) whenever changes affecting research data occur during the course of the research project.

Data Management Plans reflect research practices and the requirements of funding agencies (which may vary). In research projects conducted by consortia, two scenarios are possible depending on partner practices or funder requirements:

- the consortium prepares a joint plan that incorporates the data management practices of each partner (e.g., Horizon Europe), or
- each partner maintains its own plan (as in some calls by the Polish National Science Centre – NCN).

Open Access to Data

Open science constitutes the prevailing standard. Research data should be shared in open access unless justified limitations or other obstacles exist (e.g., related to the best interests of beneficiaries).

Use of Trusted Repositories

The use of trusted repositories is an important element of making data available to interested stakeholders. Trusted repositories provide appropriate conditions for data security, storage, and curation. They often hold certifications such as CoreTrustSeal or ISO.

Datasets stored in repositories should include metadata, including information on research funding sources, and should be assigned an appropriate open license, such as Creative Commons Public Domain Dedication (CC0) or Creative Commons Attribution (CC BY) or an equivalent license. It is also recommended that they have a Persistent Identifier (PID) such as a DOI.

The choice of repository should not only consider technical requirements and cybersecurity but also substantive considerations—namely, where the data will achieve the best visibility, citability, and potential for reuse. This is the most optimal solution from the perspective of both the researcher and the funding body.

Researchers may choose among disciplinary or thematic repositories, institutional repositories, or general repositories.

FAIR Principles (Findable, Accessible, Interoperable, Reusable)

Data should be easy to find and accessible to interested researchers, technically and interdisciplinarily interoperable, and reusable. To achieve this, it is necessary to provide detailed descriptions of data or datasets in the form of metadata and documentation, and to assign them unique, persistent identifiers (PID) (<https://www.go-fair.org/fair-principles/>).

Data Storage by Institutions

- In cases where data are stored by institutions involved in their collection, the following issues should be considered immediately after recording the data:
- access to data prior to anonymization or pseudonymization should be restricted as much as possible;
- raw, non-anonymized data should be stored for a limited period in a space isolated from anonymized/pseudonymized data (in accordance with institutional policy);

- data should be stored according to the 3-2-1 rule, and there should be procedures for regular backups; in the event of data loss or destruction, there should be a procedure for data recovery; data stored exclusively in institutional cloud spaces (not private ones) should be secure and, where possible, encrypted;
- if backup copies are stored on servers, the server should be properly secured and not publicly accessible (e.g., without Internet access).

5. Conclusion

In summary, responsible data management in research involving vulnerable groups requires a dual commitment: protecting participants from foreseeable harm while simultaneously supporting rigorous scientific research. This commitment begins with preparing an application for the Research Ethics Committee and continues through ethical engagement with participants, including providing clear and accessible information, ensuring voluntary participation, and implementing consent procedures that explicitly address potential risks, expected benefits, and circumstances in which confidentiality may be limited for protective or legal reasons (UKRI, 2025; European Commission, 2021).

It is equally important to implement rigorous data protection measures at all stages of the research project life cycle. Researchers must incorporate data protection from the research design stage onward and consistently apply it throughout implementation. This includes ensuring that the collection of personal data is necessary and proportionate, that the use of sensitive data is clearly justified, that access to data is appropriately restricted, and that data retention periods are limited to the minimum necessary for research purposes and accountability requirements (European Commission, 2021; European University Institute, 2022).

When working with vulnerable groups—where the level of risk is inherently higher—structured safeguards such as robust governance frameworks, well-designed data storage strategies, and context-appropriate protective measures are essential for maintaining compliance and minimizing the risk of disclosure, misuse, or re-identification of data (European Data Protection Board, 2021; European Commission, 2021).

Finally, good practice requires carefully balancing the principles of open science with the imperative to protect participants belonging to vulnerable groups. Although open science and data reuse can enhance transparency, reproducibility, and research impact, data sharing must never occur at the expense of participants' safety, privacy, or

dignity. Any restrictions on data sharing should be justified and proportionate and supported by strong data governance mechanisms, including controlled access procedures, comprehensive documentation, privacy-preserving techniques, and secure data storage environments (UNESCO, 2021; ALLEA, 2023). Consistent application of these principles enables research teams to comply with legal requirements and research ethics standards, maintain the trust of participants and their communities, meet contemporary standards of research integrity and responsibility, and ensure that scientific progress does not come at the expense of those most vulnerable to harm.

References

ALLEA. (2023, June 23). The European Code of Conduct for Research Integrity (2023 revised edition). <https://allea.org/code-of-conduct/>

European Commission, Directorate-General for Research and Innovation. (2021, July 5). Ethics and data protection (Horizon Europe guidance note).

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

European Data Protection Board. (2021, September 7). Legal study on the appropriate safeguards under Article 89(1) GDPR for the processing of personal data for scientific research. https://www.edpb.europa.eu/our-work-tools/our-documents/legal-study-appropriate-safeguards-under-article-891-gdpr-processing_en

European Union Agency for Cybersecurity (ENISA). (2021, January 28). Data pseudonymisation: Advanced techniques and use cases.

<https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases>

European University Institute. (2022). Guide on good data protection practice in research.

<https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

Information Commissioner's Office. (n.d.). Data protection impact assessments.

Retrieved February 4, 2026, from <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/data-protection-impact-assessments/>

UK Research and Innovation, Economic and Social Research Council. (2025, May 12). Research with potentially vulnerable people.

<https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/research-with-potentially-vulnerable-people/>

UNESCO. (2021, November 23). UNESCO recommendation on open science.

<https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>

Galica, Natalia. (2025). Introduction to Research Data Management (RDM). Presentation during the seminar MANAGEMENT OF SENSITIVE RESEARCH DATA: ANONYMIZATION, STORAGE, AND CONTROLLEDACCESS, organized within the MIGRAEDU project.

Wieczorek, Jan. (2025). Working with sensitive data: anonymization and data security. Presentation during the seminar MANAGEMENT OF SENSITIVE RESEARCH DATA: ANONYMIZATION, STORAGE, AND CONTROLLEDACCESS, organized within the MIGRAEDU project.

Wyczik, Jakub. (2025). GDPR Basics for Researchers. Presentation during the seminar MANAGEMENT OF SENSITIVE RESEARCH DATA: ANONYMIZATION, STORAGE, AND CONTROLLEDACCESS, organized within the MIGRAEDU project.