EPFL

FAST GUIDE #8

PERSONAL DATA MANAGEMENT

EPFL LIBRARY
RESEARCH DATA MANAGEMENT

v3 – 2023
DOI: 10.5281/zenodo.3327829

DEFINITIONS

Personal data is "all information relating to an identified or identifiable person" [1]

Examples: Name, Date of birth, Address, Photos, Videos, IP address, GPS coordinates, Telephone number, Credit card number, Number plate,...

Sensitive personal data is personal data "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade union membership; genetic data, biometric data processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation" [2]

FEDERAL ACT ON DATA PROTECTION (FADP) [3]

Applies to projects conducted in Switzerland, with additional laws for research involving human beings (Human Research Act) [4]
***Principles**: Good faith, Lawfulness, Proportionality, Exactitude, Security*

- **Data collected on the internet** [5] is still submitted to restrictions, even if published by the subjects
- **Hash** the identifiers if the project goals can be reached without them, and restricted access right to the **pseudomisation** key
- You need to assess the **risk of reidentification**
- **Inform the subjects** about the contact details of your unit, the purposes of your data collection, the recipients of the personal data, their right to access their personal data, and the likely consequences if they refuse to provide their personal data
- Anonymized data received from a **third party** still requires the subject to be informed of this new use
- Legal consent for subjects **under 18** years is required to collect their data
- Personal data can be **published** only if the subject consents to publication, but in no case if they are sensitive
- Guarantee these subjects' **minimal rights**: rights of access, modification, erasure

Revised Federal Act on Data Protection (FADP) in 2020
Expected to enter into force in September 2023 [8]

GENERAL DATA PROTECTION REGULATION (GDPR) [6]

Applies to projects involving personal data of subjects who are in the EU, with some derogations for scientific or statistical purposes (art.89)
***Principles**: Lawfulness, Data minimization, Accuracy, Storage limitation, Integrity, Transparency, Privacy-by-design, Confidentiality, Accountability*

- Keep a **description** of how you will implement the principles
- If data processing or storage are **outsourced**, document external services' GDPR compliance
- In the event of a **data breach**, notify the VPSI or the DSPS immediately
- **Inform subjects** of their rights to modify their data, restrict the use and withdraw their participation, as well as extensive information about data collection/processing
- Provide a **Data Protection Impact Assessment (DPIA)** [7] if the project may result in a high risk, i.e. if it involves data processed on a large scale, innovative use of data, sensitive data, vulnerable subjects, data transfers outside the EU, ...
- Any transfer of personal **data abroad** is only guaranteed for transfers to countries [8,9] whose legislation ensures an adequate level of protection
- Guarantee subjects' **minimal rights**: rights of access, rectification, portability, objection, erasure

Applicable as of May 2018

Any doubt?
Contact the EPFL Human Research Ethics Committee [10]

Credits and sources
[1] admin.ch/opc/en/classified-compilation/19920153/index.html#a3
[2] ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en /
[3] https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en /
[4] admin.ch/opc/en/classified-compilation/20061313/index.html
[5] edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/Internet_und_Computer/services-en-ligne/medias-sociaux.html
[6] https://gdpr-info.eu/
[7] ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
[8] edoeb.admin.ch/dam/edoeb/fr/dokumente/2018/staatenliste.pdf.download.pdf/20181213_Staatenliste_f.pdf
[9] ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en
[10] https://www.bj.admin.ch/bj/fr/home/staat/gesetzgebung/datenschutzstaerkung.html

Bibliothèque

Contact and info
go.epfl.ch/rdm
researchdata@epfl.ch

EPFL

FAST GUIDE #9

DATA MASKING

EPFL LIBRARY
RESEARCH DATA MANAGEMENT

v3 – 2023
DOI: 10.5281/zenodo.3327829

DEFINITION [1]

Data masking, also called data obfuscation, is the process of **hiding original data** with modified content

ADVANTAGES

Why it is worth
- ✓ Complies with law
- ✓ Makes data sharable
- ✓ Prevents data misuse
- ✓ Makes data publishable

APPLICABILITY

Tests on humans / sensitive data
- ✓ Name, identification number, location data, online identifier, etc.
- ✓ Factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity

PSEUDOANONYMIZATION

✓ FOR ACTIVE DATA
✓ REVERSIBLE

- REPLACING

  Replace data by identifiers. Store the key separately and securely.
- ENCRYPTING

  Encrypt the data and store the key securely. Appropriate for long-term preservation, not for data publishing.

≠

ANONYMIZATION

✓ FOR PUBLISHED DATA
✓ IRREVERSIBLE

- REMOVING

  Suppress data or part of the outlier records. Appropriate for processing identifiers.
- GENERALIZING

  Diminish granularity by generalizing the variables. Appropriate for data too specific or unique records.
- SHUFFLING

  Shuffle data over one / several columns without compromising their utility.
- FAKING

  Prevent the identification of specific records, adding fake data while preserving correlations.

HINT

Mitigate the identification risk, but preserve the data utility for research.

CHECK THESE TOOLS

MASK IDENTITY OR ASSESS IDENTIFICATION RISKS

- ✓ GRAASP insights [2]
- ✓ ARX Data Anonymization Tool (Java) [3]
- ✓ Amnesia [4]
- ✓ ARGUS (Java) [5]
- ✓ sdcMicro (R) [6]
- ✓ Differential privacy queries (SQL) [7]
- ✓ Faker (Python) [8]
- ✓ OpenPseudonymiser [9]
- ✓ AES Crypt [10]

Do you deal with personal data?

Check the **Fast Guide #8: PERSONAL DATA MANAGEMENT** [12]

- EPFL Research Ethics [13]
- Federal Act on Data Protection (FADP) [14]
- Human Research Act (HRA) [15]
- General Data Protection Regulation (GDPR) [16]

Credits and sources
[1] en.wikipedia.org/wiki/Data_masking
[2] insights.graasp.org
[3] arx.deidentifier.org
[4] amnesia.openaire.eu/
[5] qosient.com/argus/ anonymization.shtml
[6] https://cran.r-project.org/web/packages/sdcMicro/index.html
[7] github.com/uber/sql-differential-privacy
[8] hfaker.readthedocs.io/en/master/
[9] openpseudonymiser.org
[10] www.aescrypt.com
[12] go.epfl.ch/rdm-fastguide08
[14] https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en
[15] admin.ch/opc/en/classified-compilation/20061313/index.html
[16] https://gdpr-info.eu

Bibliothèque

Contact and info
go.epfl.ch/rdm
researchdata@epfl.ch