

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



ANONYMIZATION



- ✓ FOR ACTIVE DATA
- ✓ REVERSIBLE



- ✓ FOR PUBLISHED DATA
- ✓ IRREVERSIBLE

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.

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.

UTILITY PROTECTION



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 [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](https://en.wikipedia.org/wiki/Data_masking)
- [2] [insights.graasp.org](https://insights.graasp.org)
- [3] [arx.deidentifier.org](https://arx.deidentifier.org)
- [4] [amnesia.openaire.eu/](https://amnesia.openaire.eu/)
- [5] [qosient.com/argus/anonymization.shtml](https://qosient.com/argus/anonymization.shtml)
- [6] <https://cran.r-project.org/web/packages/sdcMicro/index.html>
- [7] [github.com/uber/sql-differential-privacy](https://github.com/uber/sql-differential-privacy)
- [8] [hfaker.readthedocs.io/en/master/](https://hfaker.readthedocs.io/en/master/)
- [9] [openpseudonymiser.org](https://openpseudonymiser.org)
- [10] [www.aescrypt.com](https://www.aescrypt.com)
- [11] [go.epfl.ch/rdm-fastguide08](https://go.epfl.ch/rdm-fastguide08)
- [14] [https://www.fedlex.admin.ch/eli/cc/1993/1945\\_1945\\_1945/en](https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en)
- [15] [admin.ch/opc/en/classified-compilation/20061313/index.html](https://admin.ch/opc/en/classified-compilation/20061313/index.html)
- [16] <https://gdpr-info.eu>