DEEPHEALTH

Hackathon - Course 1

Hackathon presentation

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Index

1. Introduction and Course Index
2. Pseudonymisation vs. Anonymisation
   a. Definition
   b. Basic Principles
   c. Main Risks
   d. Anonymisation techniques
   e. Pseudonymisation techniques
3. Clinical Data and Relevance of De-Identification
   a. Medical Information Systems
   b. MIDS
   c. Types of anonymisable clinical data
4. GDPR on Clinical Data
   a. GDPR
   b. Data ownership
   c. Clinical Data and GDPR
5. BIMCV

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

DEEPHEALTH

Introduction and Course Index

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Introduction

- How all the methodology associated with anonymisation methods and their implementation have been developed.

- Provide guidelines and best practices for the medical imaging community

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Course Index

1. Hackathon presentation
2. De-identification of Radiological reports
3. De-identification of DICOM metadata
4. De-identification of Biomedical Images
5. Workshops

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

DEEPHEALTH

Pseudonymisation vs. Anonymisation

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Definitions

De-identification is the process used to prevent someone's personal identity from being revealed.

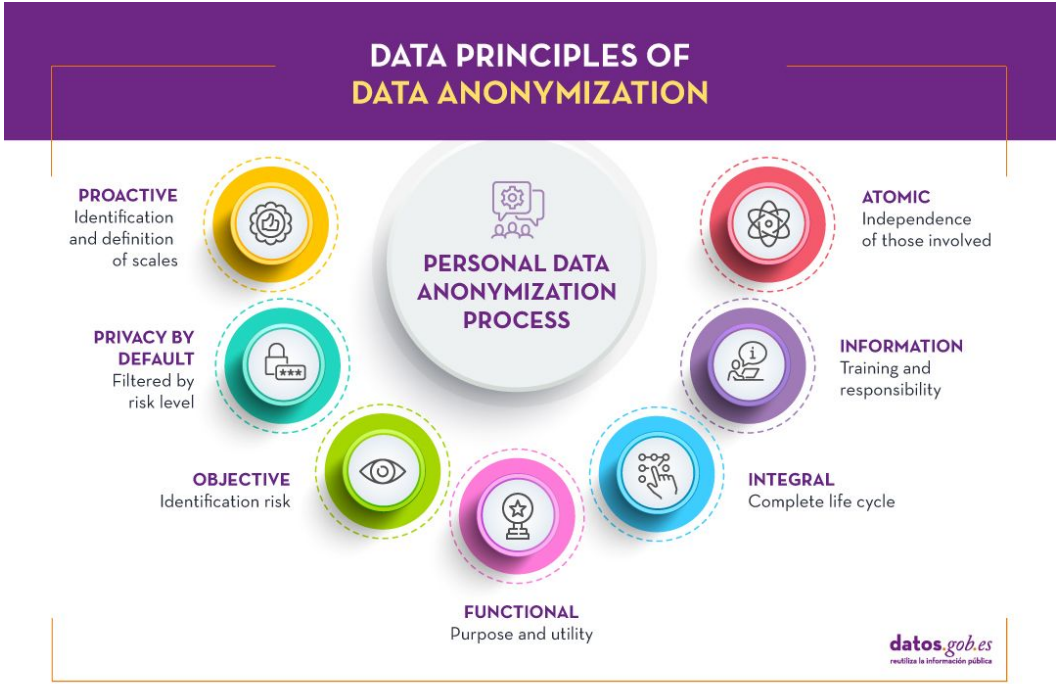
The **anonymisation** of data is an irreversible process. Anonymised data cannot be linked to the original subject it identifies.

Pseudonymisation limits the traceability between the data and the original subject it identifies. It can be reversed.

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Basic principles

- Proactive
- Privacy by default
- Objective
- Functional
- Integral
- Information
- Atomic



https://datos.gob.es/en/blog/importance-anonymization-and-data-privacy

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Main risks

- **Singling out**: risk of extracting attributes that allow an individual to be identified.

- **Linkability**: risk of linking at least two attributes to the same individual or group, in one or more data sets.

- **Inference**: risk of deducing the value of a critical attribute from other attributes.

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Anonymisation

Methods

- Randomisation
  - noise addition
  - permutation
  - differential privacy

- Generalisation
  - aggregation and anonymity "k"
  - diversity "l" / proximity "t"

- Encoding
  - Hash algorithms with secret key and key erasure
  - homomorphic encryption
  - time stamp

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Pseudonymisation

Methods

The processing of personal data such that it can no longer be attributed to a specific data subject without using additional information.

- Hash algorithms
- Secret key encryption
- Token decomposition

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

DEEPHEALTH

Clinical data and relevance of de-identification

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Medical Information Systems

HIPAA

The Health Insurance Portability and Accountability Act (**HIPAA**) proposes a set of good practices to protect:

- confidentiality
- integrity
- availability

of information in the health sector.

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Medical Information Systems

Standards

Standards for the development and management of hospital information:

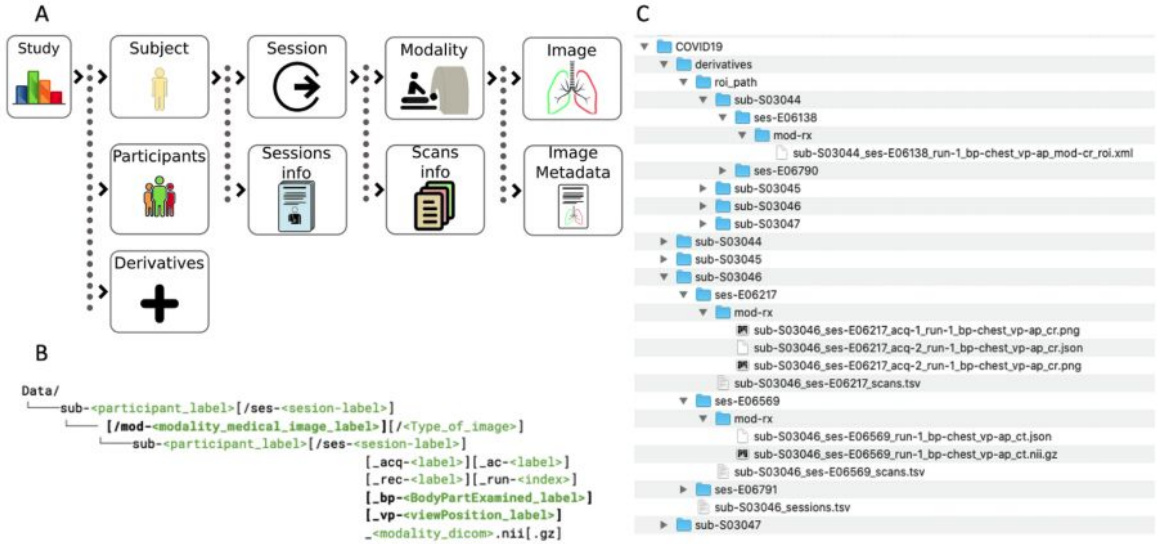
- **HL7** (Health Level Seven Inc.) → standardise the exchange of medical information among applications

- **DICOM** (Digital Imaging and Communications in Medicine) → ensure the interoperability among heterogeneous medical imaging equipment and systems

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

MIDS

Medical Imaging Data Structure (**MIDS**) is the methodology proposed by FISABIO to standardise the organization and management of medical imaging data.



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Types of anonymisable clinical data

Radiological reports

- Traditionally de-identified through pattern matching and regular expressions
- Natural Language Processing (NLP) de-identification as an alternative

DICOM meta information

- Data De-identification Guidelines and Protocol
- **Smart-Upload:** FISABIO's open-source DICOM de-identification software tool

Pixel data and graphic information

- Facial information
- Text annotations superimposed to the image

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

DEEPHEALTH

GDPR on Clinical Data

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

GDPR

General Data Protection Regulation

- It is applied to the processing of personal data
- It is applied in the form of national laws.
- Possibility of varying interpretations and legal implementations

Personal data must be collected, processed and shared under a lawful basis of the GDPR:

- Informed consent
- Public interest
- Legitimate interest

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Data ownership

- The GDPR does not explicitly define or assign data ownership

- The subject has to give consent for:
  - collection
  - processing
  - sharing

  of their data

- This consent can be revoked at any time

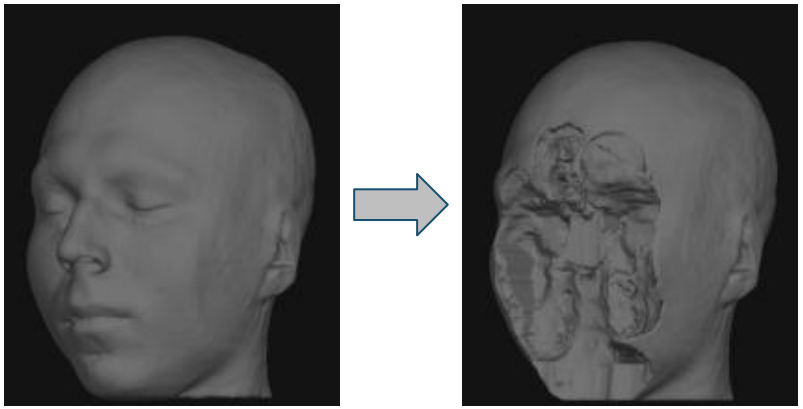
The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Clinical Data and GDPR

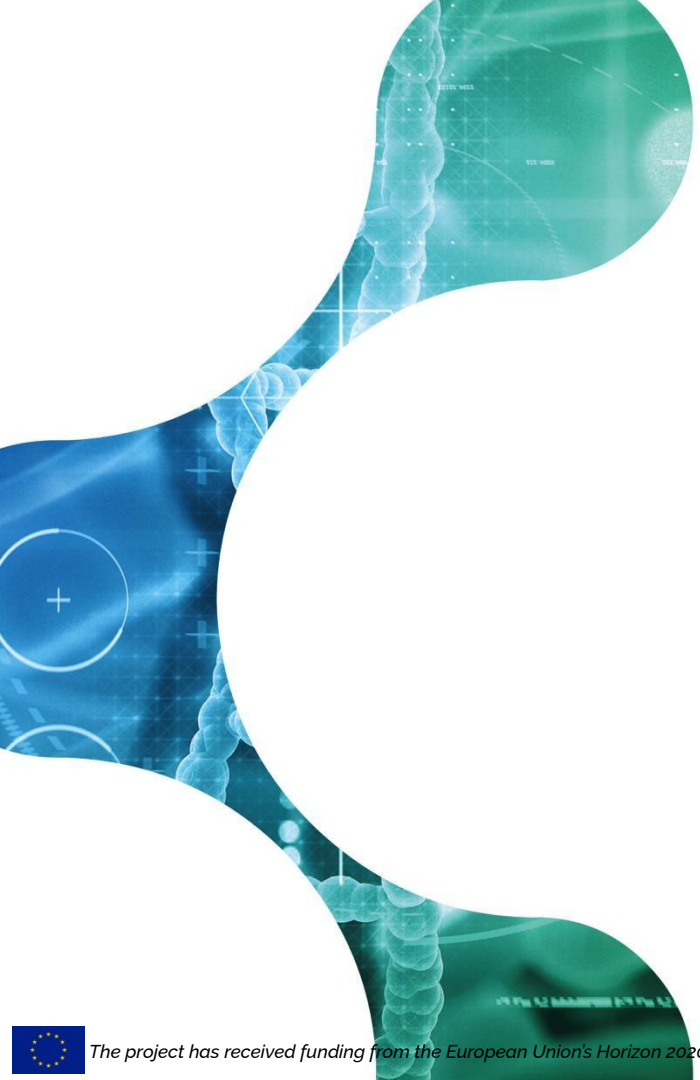
What constitutes anonymous data when it comes to clinical data?

e.g., Brain MRI

- Are they anonymous?
  → Facial features could be reconstructed
- How can they be anonymised?
  → Defacing



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.



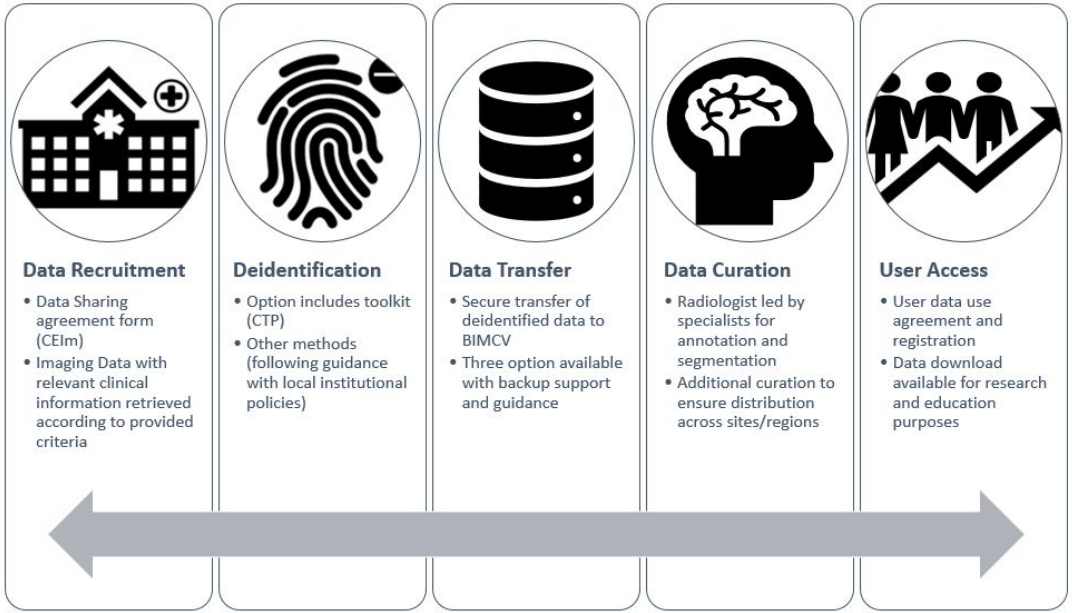
DEEPHEALTH

BIMCV

Medical Imaging Data Bank of Valencia Region

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

BIMCV life cycle



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

DEEPHEALTH

Thank you for your attention!

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Index

Rocher, L., et al. Estimating the success of re-identifications in incomplete datasets using generative models.
Nat. Commun.10, 1–9 (2019).

Yeh, F.-C. et al. Quantifying differences and similarities in whole-brain white matter architecture using local
connectome fingerprints. PLoS Comput. Biol. 12, e1005203 (2016).

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.