SCTO PLATFORMS

25 June 2025 | 12.00–13.00 | online seminar

Facts and pitfalls of observational studies

De-identification: Definition, types, relevance and operational approaches

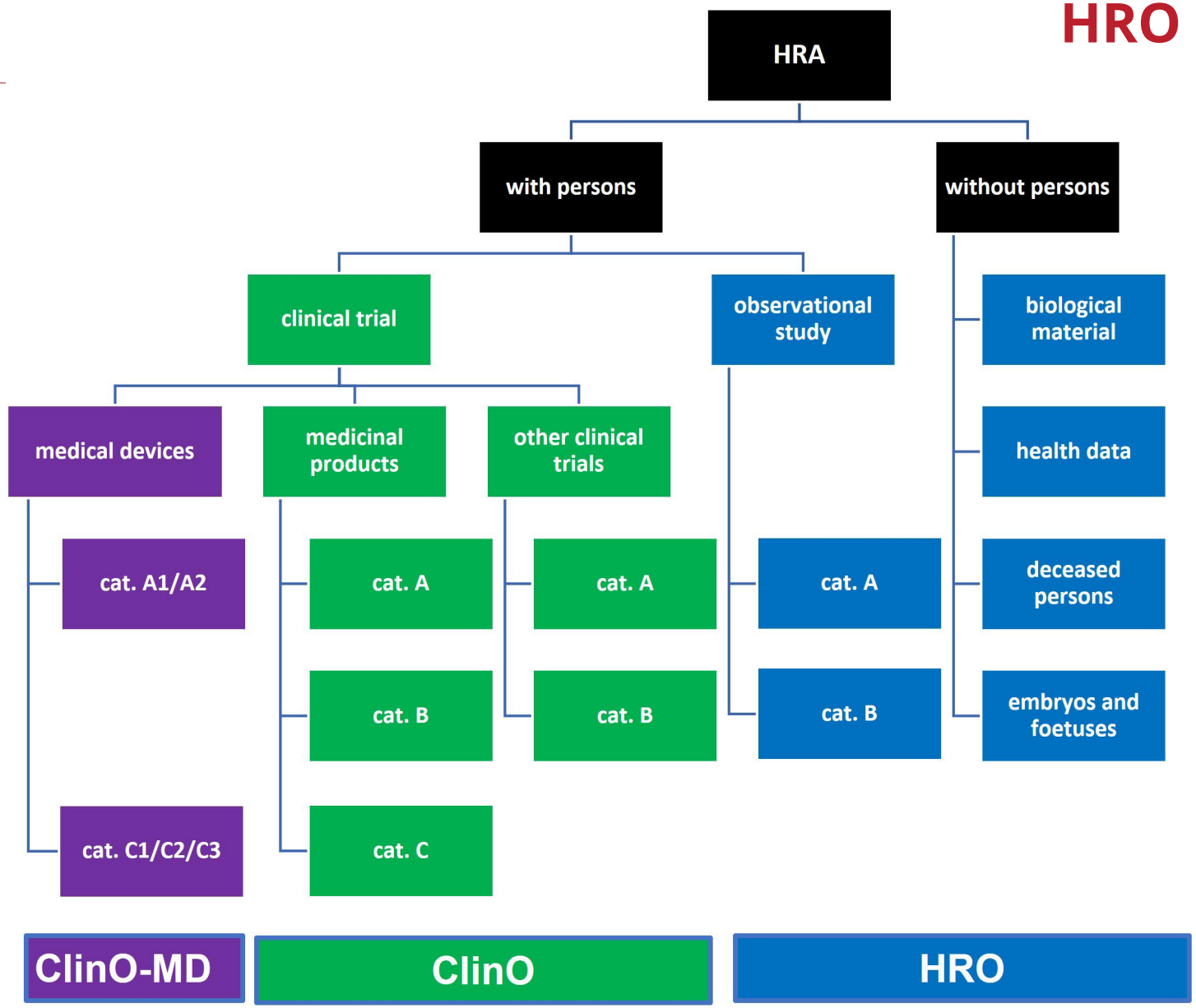
This two-part session highlights the principles and relevance of de-identification of health-related personal data in human research and provides insights into the operational practices of proper de-identification.

Registration and more information:

sctoplatforms.ch/hro-session-de-identification.ch

SCTO PLATFORMS

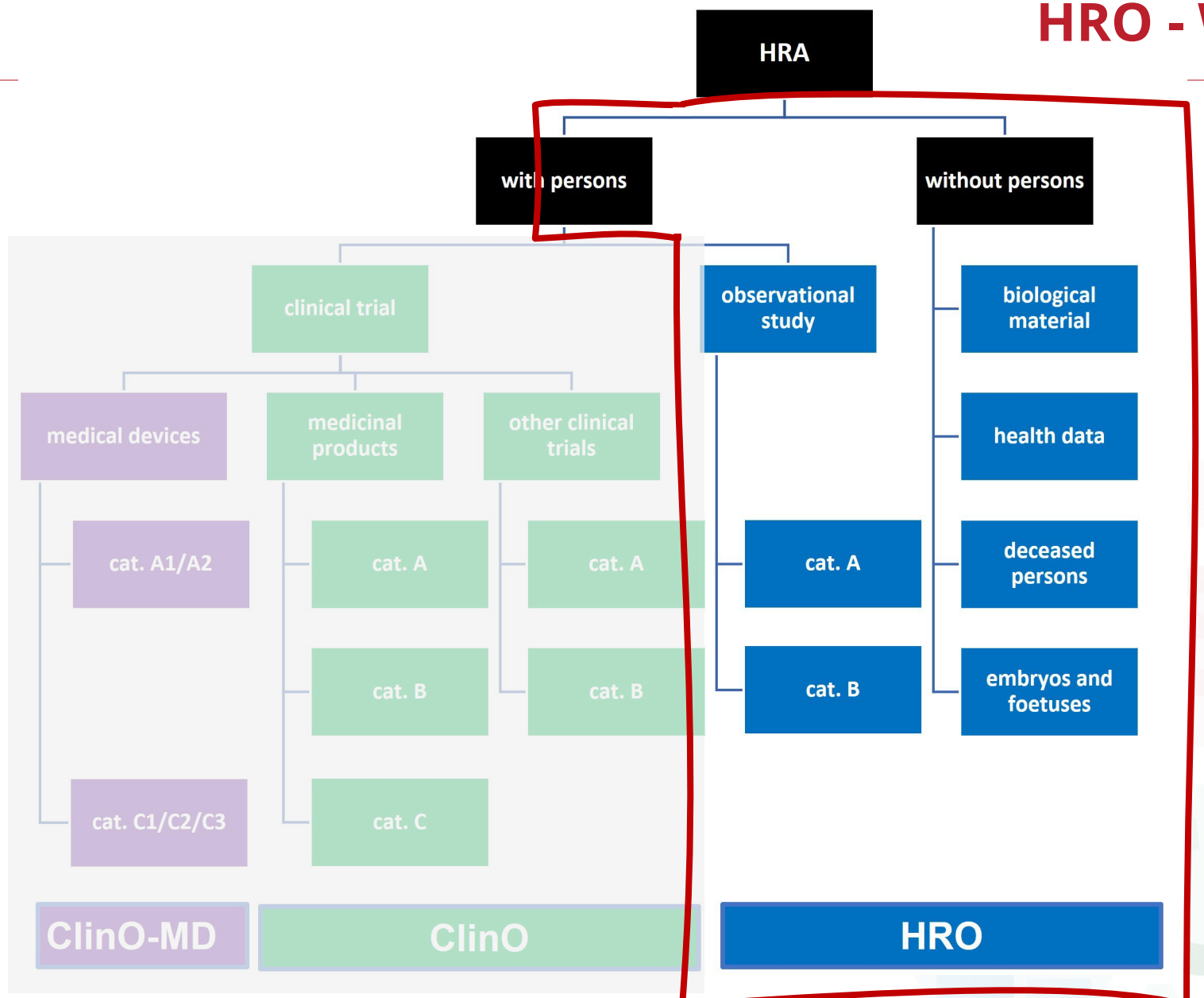
HRO - Where are we?



according to Ch. Seiler, EC Bern

SCTO PLATFORMS

HRO - Where are we?



according to Ch. Seiler, EC Bern

SCTO PLATFORMS

General information

- **Questions**:
  - **during presentation**: in the chat mentioning the part (I or II) of the session that it refers to (→ for Q&A session at the end)

- Presentation recorded

- Video, slides & Q&A provided after the session

- Feedback poll at end → please fill in!

- HRO lunch project team:
  - Claudia Fila (CTC Zurich)
  - Antoine Poncet (HUG)
  - Verena Golz (DKF Basel)

SCTO PLATFORMS

Presentation

Dr. jur. Thomas Gruberski

Member of the Ethics Committee of Northwest and Central Switzerland (EKNZ), Department of Clinical Research (DKF) Basel

thomastadeusz.gruberski@usb.ch

Matthias Joos, M.Sc.

Data Solution Engineer
Department of Data and Platform Services, University Hospital Zurich (USZ)

matthias.joos@usz.ch

SCTO PLATFORMS

HRO Lunch Session - Part I

Definition, types and relevance

Dr. jur. Thomas Gruberski

Universitätsspital Basel

Generalsekretariat

Rechtsdienst

De-identification - definition, types, relevance and operational approaches

The legal perspective

Dr. iur. Thomas Gruberski

A question to start with…

- (BTW The answers are being collected anonymously.)

Various applicable acts

- Preliminary Remark: It‘s not rocket science - it‘s worse.

- **Various Acts**
  - Federal Data Protection Act (new version enacted on September 1st 2023):
  - Applies for federal authorities & private organisations (such as pharma companies)
  - Cantonal Data Protection Acts, e.g. IDG BS: Applies for USB

  - Pro memoria: GDPR (“DSGVO”)

The applicable acts and their relation to each other

- **HRA plus its ordinances: “lex specialis”: Takes precedence over the Data Protection Acts.**
- If no research (in terms of the HRA [see next slide]): Data Protection Acts still applicable.
- Handling with anonymised data: Data Protection law no longer applicable.

HRA – scope of application

[1] This Act **applies** to research concerning human diseases and concerning the structure and function of the human body, which involves:

a. persons, deceased persons, embryos and foetuses;

d. biological material;

e. health-related personal data

[2] It does **not apply** to research which involves:

b. anonymised biological material;

c. anonymously collected or anonymised health-related data

Important semantic issue

- HRA & ordinances speak of “Coded” Data (“verschlüsselte Daten”).
- What they mean by that (and what has prevailed): **Pseudonymised** Data
- Reason: Coding = alteration of information by using an algorithm with the aim of making it unreadable to unauthorised persons (cryptographic encryption)

Anonymisation (Clause 25 HRO) - I

- For the anonymisation of biological material and health-related personal data, any association with a specific person must be rendered (1) **impossible** or (2) eliminated in such a way as to allow this association to be re-established **only with disproportionate effort**.

Anonymisation (Clause 25 HRO) - II

- Anonymisation must be effected using a method based on the **current state of the art**.

- In particular, items of data which, individually or in combination, allow **the association with a specific person to be re-established**, such as the first name, surname, address, date of birth or unique identification numbers, must be **deleted or modified**.

Anonymisation (Clause 25 HRO) - III

- The method used for anonymisation must be documented, including a description of the residual risk of re-identification. (BTW: why…?)

Coding (Clause 26 HRO) - I

- Biological material and health-related personal data are considered to be correctly coded in accordance with Article 32 paragraph 2 and Article 33 paragraph 2 HRA if, without access to the key or to the source data, it is only possible with disproportionate effort to link the biological material or the health–related data to a specific person.

Coding (Clause 26 HRO) - II

- Coding must be effected using a method based on the current state of the art. The key must be stored separately from the biological material or personal data and in accordance with the principles of Article 5 paragraph 1, by a person or organisational unit to be designated in the application, **not involved in the research project**.

- (PS the EC reviews if the coding process is correct and secure [Clause 34 I c HRO].)

Coding (Clause 26 HRO, old version) - III

- Biological material and health-related personal data are considered to be correctly coded in accordance with Article 32 paragraph 2 and Article 33 paragraph 2 HRA **if, from the perspective of a person who lacks access to the key, they are to be characterised as anonymised.**

Conditions for breaking the code (Clause 27 HRO)

For coded biological material and coded health-related personal data, the code may only be broken if:

- breaking the code is necessary to avert an immediate risk to the health of the person concerned;
- a legal basis exists for breaking the code; or
- breaking the code is necessary to guarantee the rights of the person concerned, and in particular the right to revoke consent.

User’s manual for coding

- Take the list containing the relevant personal data («Klardaten»).
- Remove all the identifying factors (name, birth day, address, email address, patient ID, etc.) and replace them e.g. by a number.
- Create a separate document which contains the link between the number and the person in question.
- This «separate document» is the key.

Take home message (and maybe opening the discussion)

- **Impossibility** of Re-Identification = 100% - not required by law (!)
- “Disproportionate effort” is sufficient.
- Examples:
  - Hacking into CIA’s database?
  - Matching with…..? (new data bases emerge every day.)

BTW: Joke of the day - coding of biological material

- Interesting citation about “hiding the key”:
- This would mean to *lock up* the donor of the biological material in question…

Questions or remarks ?

SCTO PLATFORMS

HRO Lunch Session - Part II

Operational approaches

Matthias Joos, M.Sc.

Data Deidentification Methods at USZ
Insights into Operational Approaches / Practices

HRO Lunch Session 25.6.2025

Matthias Joos – Data Solution Engineer, Data & Platform Services

Identifying data

Data and samples can be identified as is.

The dataset contains identifying data like patient id, case id, name, address, hospital id, ahv number, birthdate, rare diseases etc.

The majority of clinical data.

Identification possible

Pseudonymized data

Access to keyless only for persons not involved in the research project (Art. 26 HRO).

Reidentification is only allowed given an immediate health threat of the person concerned, in case of a legal imperative arbitrament or if the reidentification is required to ensure the legal rights of the person concerned, in particular an act of revocation. (Art. 27 HRO)

Name	Hospital-Id	Pseudonym
Anna Muster	112345|USZ	527D2FA6-184A-B9CA-D98A-26FF-B835C20C4DF2

Reidentification possible

Anonymized data

Irreversible deletion/redaction of all identifying data and combinations thereof, which allow the reestablishment of a reference to a person without disproportionate effort. Art. 25 HRO)

The feasibility of anonymization decreases with growing number of parameters.

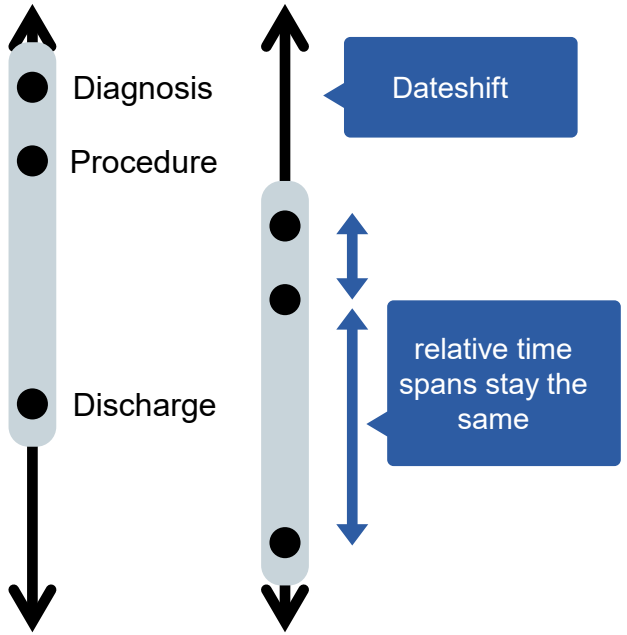
Reidentification not possible

Human Research Act HRA

National and Cantonal Data Protection Laws (EU: DSGVO)

Structured Data

- Generation of **project specific research IDs** for patients / cases or other identifiers from the source system
- Patient- and project-specific random shift of dates / datetimes
  - **preserve longitudinal consistency**
  - different rulesets (e.g. +-365 days, +- 30 days, …)



Identifying data

Name **Anna Muster**

Hospital-Nr **12345**

Admission Date **11.01.2010**

AHV-Number **756.1234.5678.91**

Identification possible

deidentification process

delete

pseudonym

date shift

delete

Pseudonymized data

Name **-**

Hospital-Nr **FXBSJDIW**

Admission Date **30.03.2010**

AHV-Number **-**

keylist

Name	Hospital-Id	Pseudonym
Anna Muster	12345	FXBSJDIW

Reidentification possible

Anonymized data

Name **-**

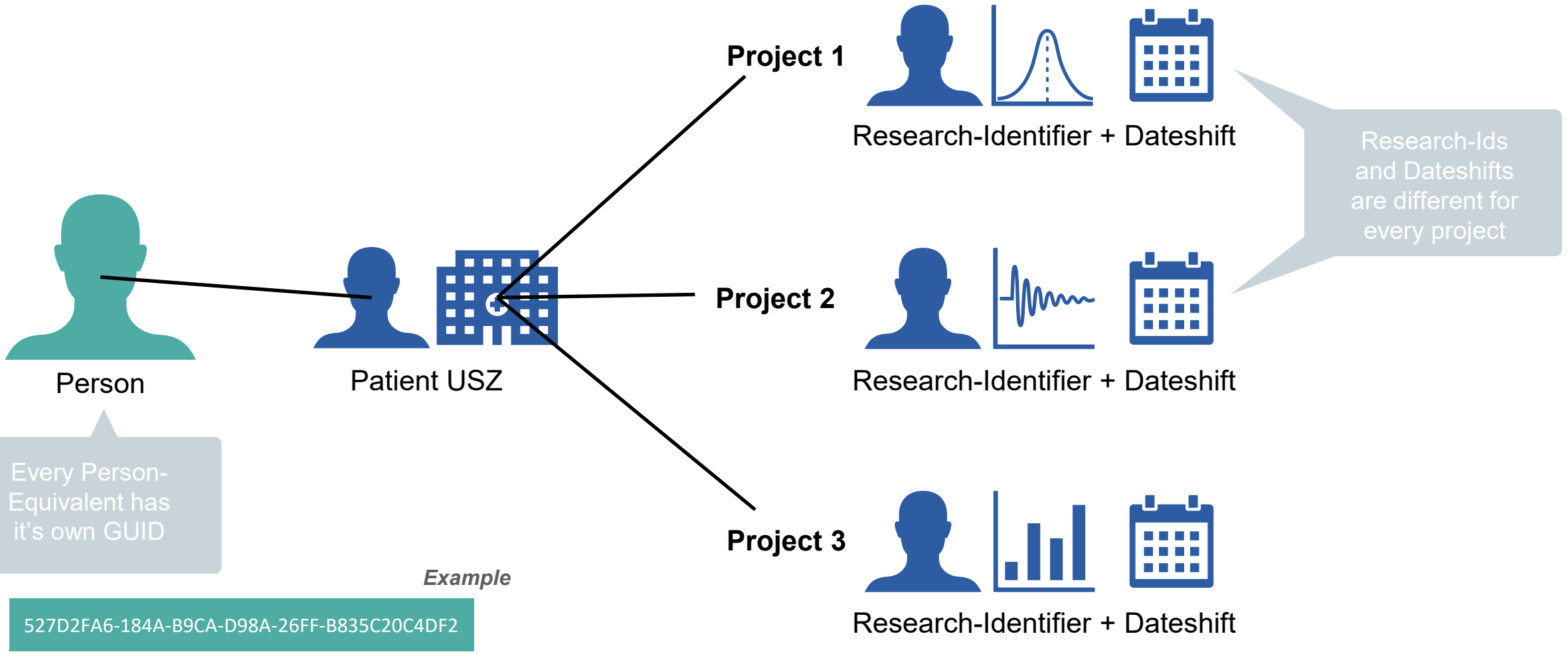
Hospital-Nr **-**

Admission Date **-**

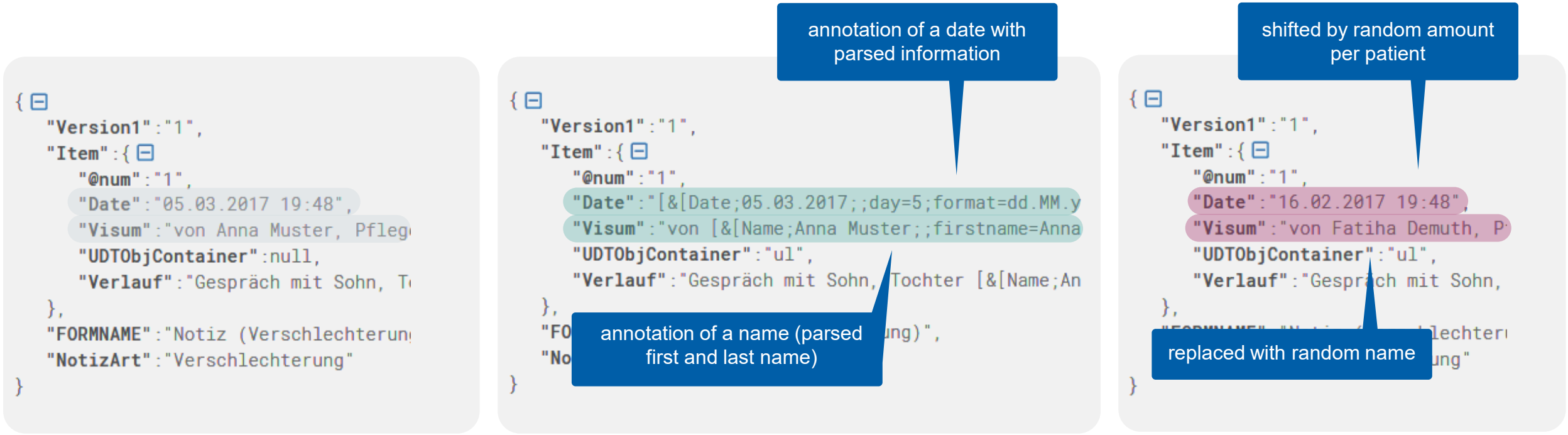
AHV-Number **-**

no keylist

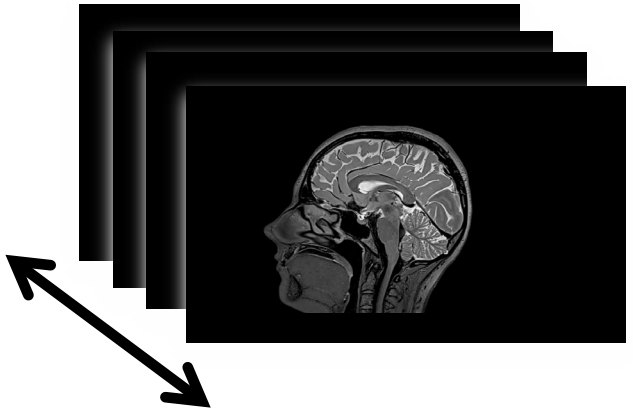
Reidentification not possible


Project 1
Research-Identifier + Dateshift
Research-Ids and Dateshifts are different for every project
Project 2
Research-Identifier + Dateshift
Person
Patient USZ
Every Person-Equivalent has it's own GUID
Project 3
Research-Identifier + Dateshift
Example
527D2FA6-184A-B9CA-D98A-26FF-B835C20C4DF2

Unstructured Text Reports

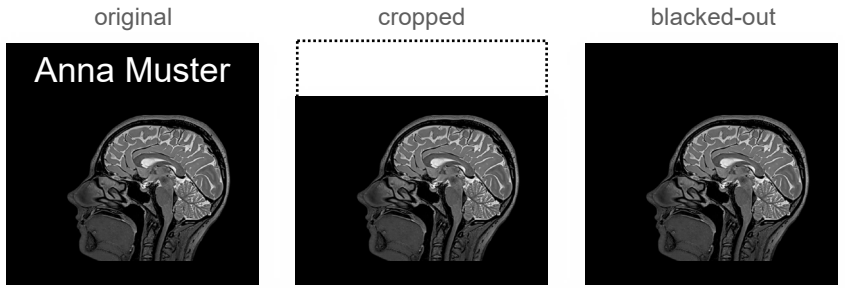


Dicom Images



Deidentification of DICOM metadata and pixel data required!

- **Dicom Tags (metadata)**
  - Generation of project specific research IDs for patients
  - Patient-specific random shift of dates / datetimes
  - Generation of project specific **research UIDs for studies, series, images, frames**
- **Pixel Data**
  - **Cut-Off** configuration for burnt-in annotations
  - **Black-Out** pixel ranges

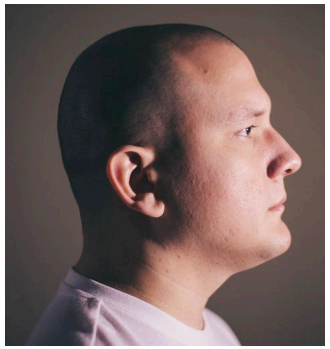


Manipulating Metadata is not enough!

Cranial MR Images

Virtual Photo from MRI

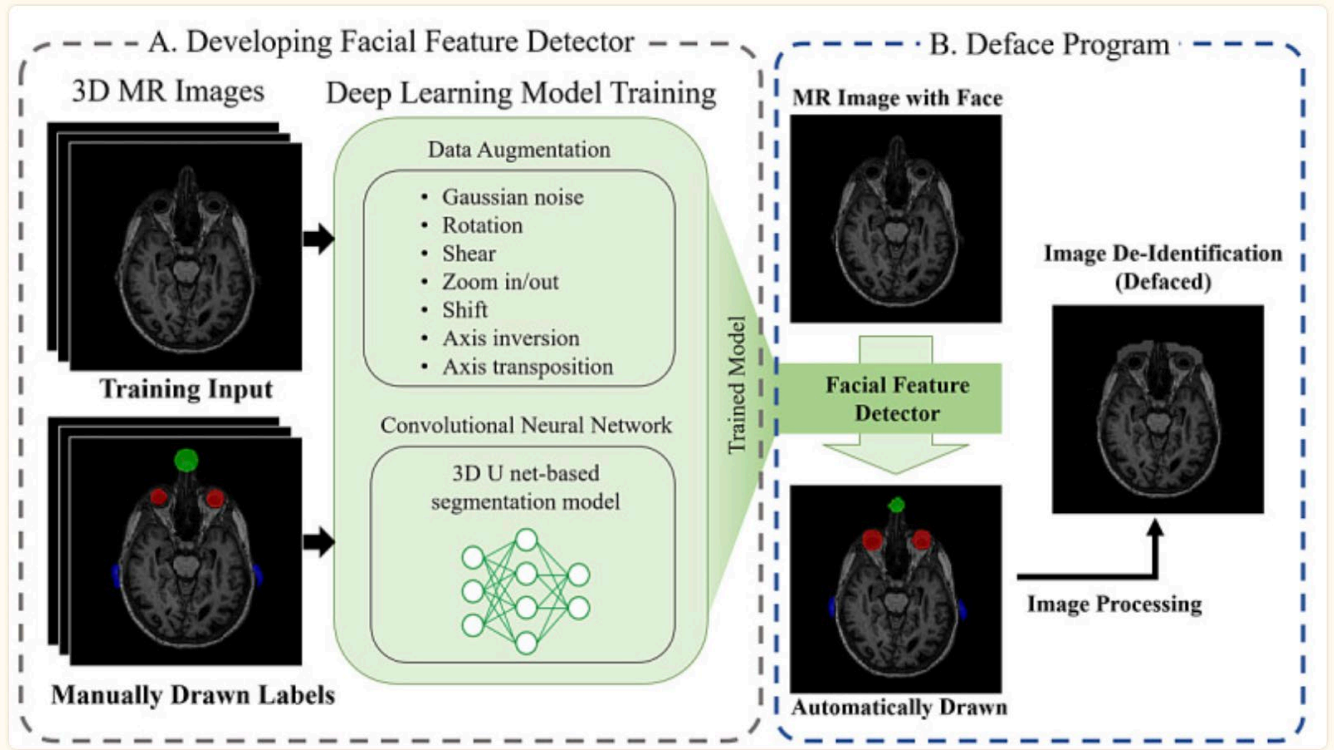
Public Photo

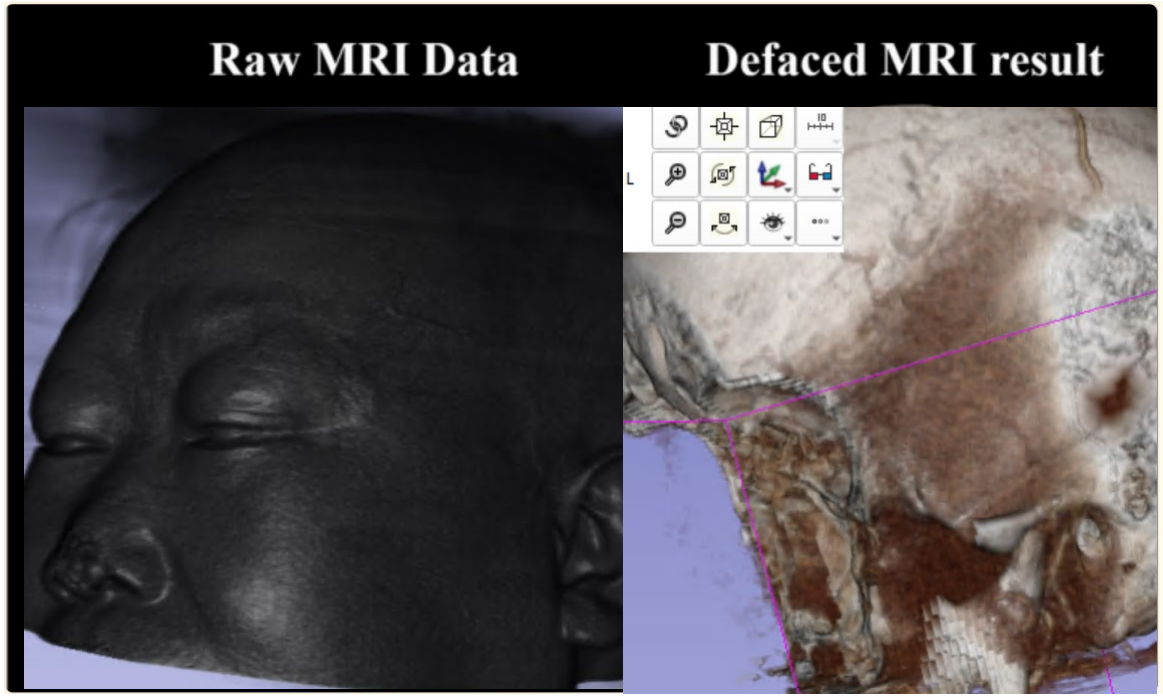


3D face reconstruction (volume rendering)

face recognition

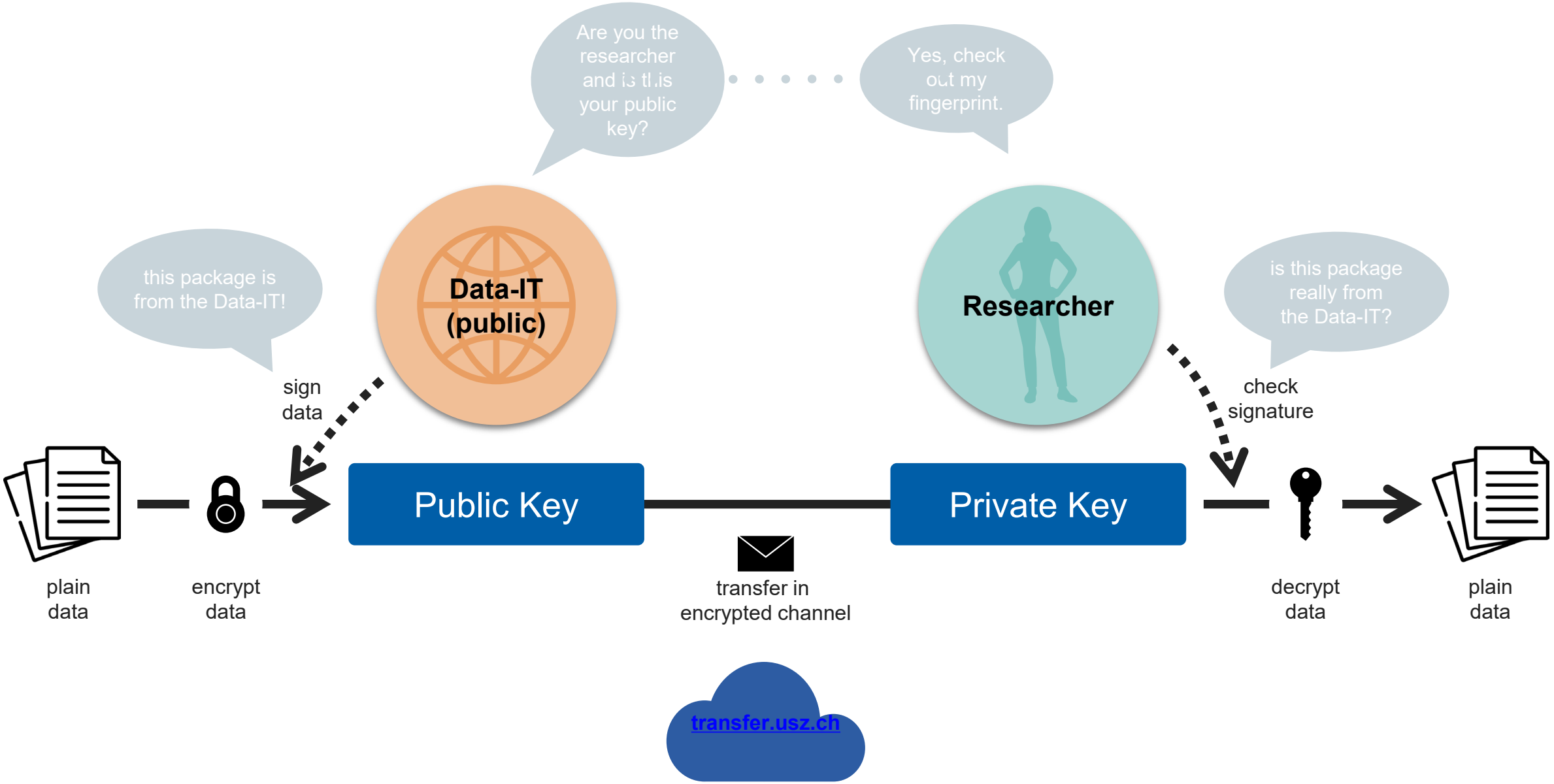
Defacing Cranial MR Images





Process of (A) developing the facial feature detector, which is a deep learning model that can detect the eyes, nose, and ears in 3-dimensional (3D) magnetic resonance (MR) images, and (B) distorting the facial features in nonanonymized cranial MR images.

Source: J Med Internet Res. 2020 Dec; 22(12): e22739. „De-Identification of Facial Features in Magnetic Resonance Images: Software Development Using Deep Learning Technology“ (Yeon Uk Jeong, BSc, PharmD, Soyoung Yoo, PhD, Young-Hak Kim, MD, PhD and Woo Hyun Shim, PhD)


Are you the researcher and is this your public key?
Yes, check out my fingerprint.
this package is from the Data-IT!
Data-IT (public)
Researcher
is this package really from the Data-IT?
sign data
check signature
Public Key
Private Key
plain data
encrypt data
transfer in encrypted channel
decrypt data
plain data
transfer.usz.ch

Welcome to the Privacy Toolbox!

Our mission is to advance the field of biomedical data research by harnessing the expertise of the SPHN DeID task force and translating their recommendations into a dynamic, adaptable platform. Our goal is to revolutionize the process of risk assessment and de-identification for biomedical datasets, streamlining it for research purposes. In pursuit of this objective, we are committed to developing an automated de-identification tool that not only ensures transparency but also provides a clear understanding of risk levels. This innovation promises to be a valuable asset for researchers, regulatory authorities, and Data Protection Officers (DPOs) alike. Join us on this exciting journey towards enhanced biomedical data privacy and research efficiency.

Qualitative Risk Assessment

Analyze and assess risk levels in your data.

Quantitative Risk Assessment

Placeholder text

Rule-Based De-identification

Placeholder text

Text DeID

Automate text de-identification with ease.

Synthetic Data Generation

Generate synthetic data for research purposes.

SCTO PLATFORMS

SPHN: Resources for health data de-identification

- SPHN supports researchers in **de-identifying personal health data** in line with Swiss legal requirements
- Resources include:
  - An expert **Guidance Paper** outlining a sound data de-identification methodology
  - A **practical Excel Risk Assessment Template** to help implement the methodology in practice

Access the Guidance Paper and Excel Template here:

Swiss Personalized Health Network

SPHN

Q&A session – questions?

Part1: Thomas Gruberski

Part 2: Matthias Joos

Thank you for participating!

Further questions to:

thomastadeusz.gruberski@usb.ch

matthias.joos@usz.ch

Verena.Golz@usb.ch

Information on
HRO lunch session 3
- 1 October 2025 -
will be provided as soon as possible under:

Don't miss out on valuable training opportunities — subscribe to the SCTO newsletter today:

www.scto.ch/en/news/newsletter.ch

Explore and download free tools designed for the clinical research community:

www.sctoplatforms.ch