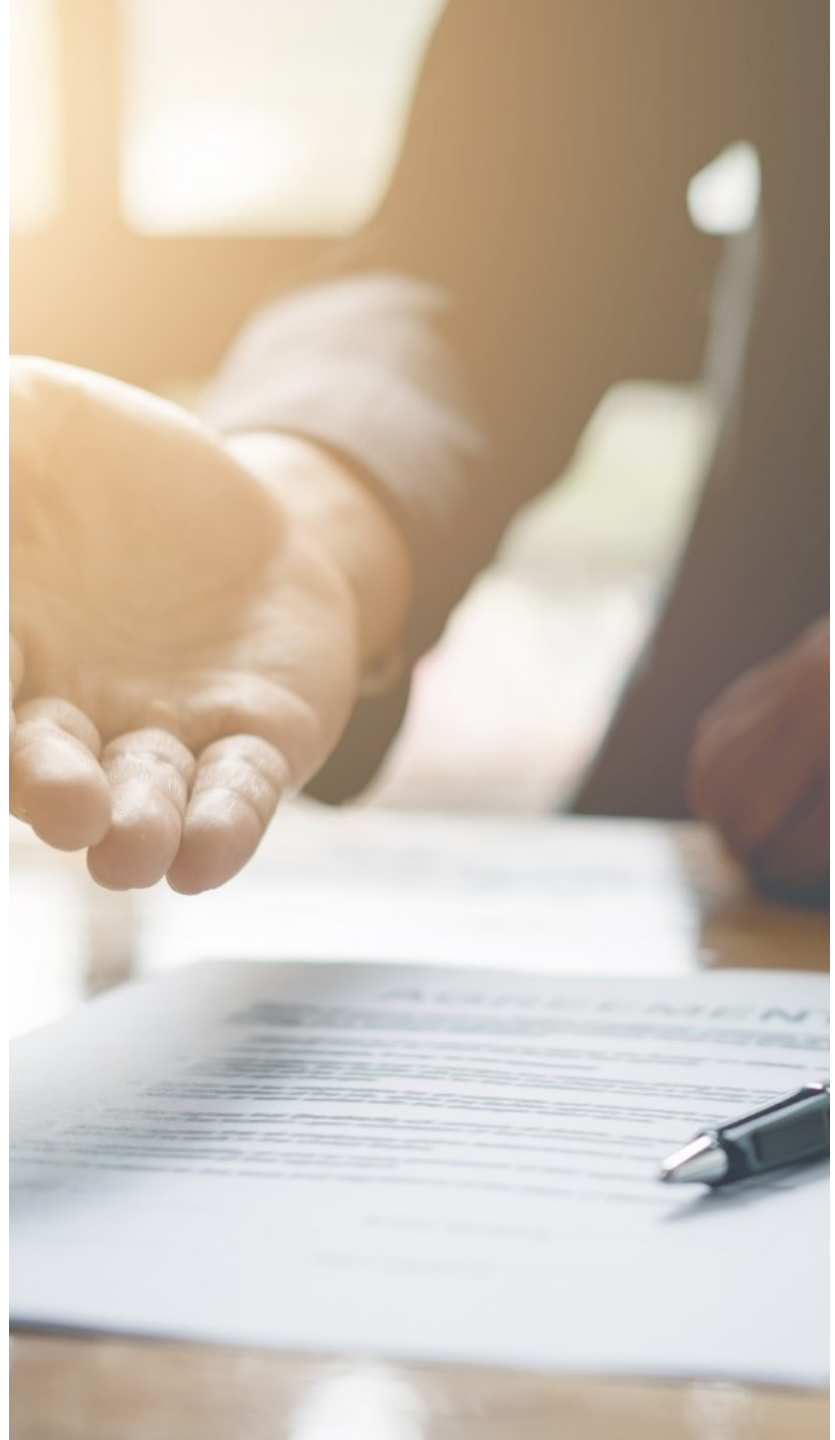
# *How GDPR fosters pseudonymisation in academic research.*

# *The perspective of a university hospital DPO.*

Dr. Griet Verhenneman – 9 December 2021 – IPEN webinar 2021: “Pseudonymous data: processing personal data while mitigating risks”

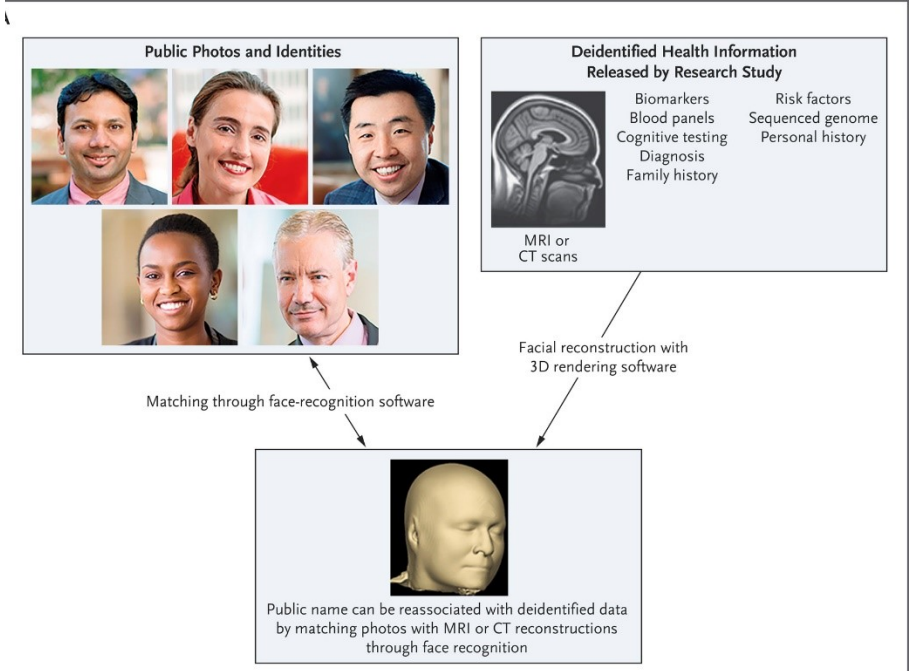


The NEW ENGLAND JOURNAL of MEDICINE

CORRESPONDENCE

# Identification of Anonymous MRI Research Participants with Face-Recognition Software

**TO THE EDITOR:** Public sharing of research data is being widely promoted. Medical image files contain "metadata" such as the name of the participant, the date of the scan, and the identification number. Such data are typically removed (deidentified) before data sharing, but images of the face in magnetic resonance imaging (MRI) MRI scans, we recruited 84 volunteers between the ages of 34 and 89 years, stratified according to sex and decade of age, and photographed each participant's face from five slightly varying angles. Each participant had undergone MRI of the head (three-dimensional fluid-attenuated inversion recovery [FLAIR] sequence, conducted




See: Swarz e.a., "Identification of Anonymous MRI Research Participants with Face-Recognition Software", *NEJM*, 2019, 1684-1689.

*SHE REALLY LIKED THAT SHIRT —*

# Masked arsonist might've gotten away with it if she hadn't left Etsy review

Woman who burned two police cars IDed by tattoo and Etsy review of her T-shirt.

JON BRODKIN - 6/18/2020, 6:48 PM


FBI

Enlarge / Instagram photo of a masked woman, identified by the FBI as Lore-Elisabeth Blumenthal, on May 30, 2020 in Philadelphia.

*"It is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data"*

Article 29 WP, Opinion 05/2014 on Anonymisation Techniques

47. It should be taken into account that anonymisation of personal data can be difficult to achieve (and upheld) due also to ongoing advancements in available technological means, and progress made in the field of re-identification. For this reason, the anonymisation of personal data should be approached with caution in the context of scientific research. Those parties which consider that they are using anonymous information in research should be in a position to satisfy themselves – and when questioned also the competent SA - on an ongoing basis that this continues to be the case, and that they have not inadvertently become data controllers of personal data for the purposes of the Regulation.

EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research Adopted on 2 February 2021

[Translated] "***Be aware: data are only (sufficiently) anonymous when also in combination with other data (including those held by other parties) they cannot lead to re-identification (e.g. IP addresses are always personal data because with the help of telecom operators one can be re-identified)***"

Recommendation eGezondheid GDPR and apps, 2020

(https://www.ehealth.fgov.be/nl/egezondheid/task-force-data-technology-against-corona/aanbevelingen-op-het-vlak-van-naleving-van-de-avg-door-apps)

While anonymisation techniques are crucial, data are rarely anonymous.

We are pseudonymising rather than anonymising.

# Chapter 2: How do we ensure anonymisation is effective?

## Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance

October 2021

ico.
Information Commissioner's Office

# Policy for secondary use of (pseudonymised) personal data

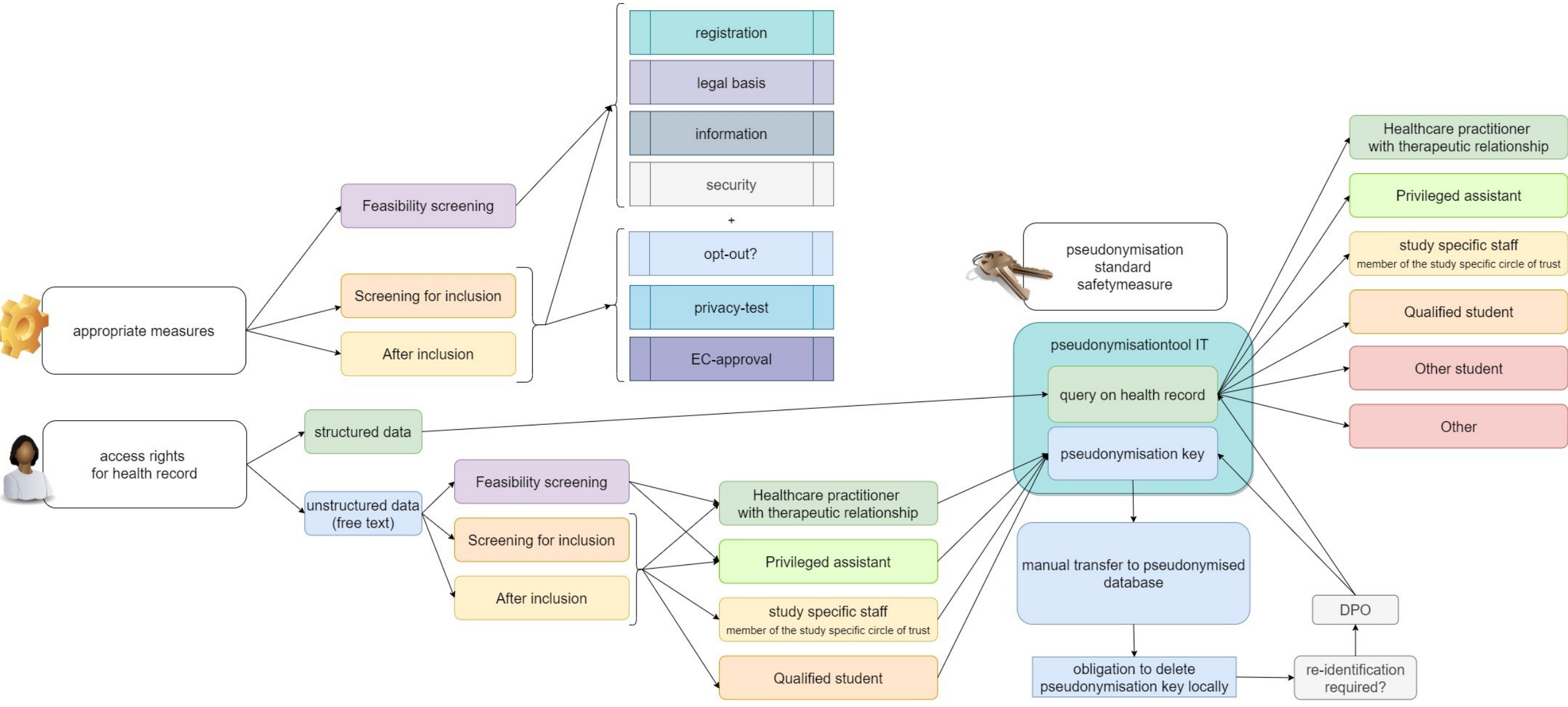
I. Conditions for secondary use:

1. Registration
2. Privacy assessment
3. Legal basis
4. A right to opt-out?
5. Right to information
6. Security

II. Direct access to medical records (= non pseudonymised) is limited

III. Pseudonymisation in application of need to know, not nice to know principle

# Policy for secondary use of (pseudonymised) personal data



# Four risk levels

| Risk level 0 | Risk level 1 | Risk level 2 | Risk level 3 |
| --- | --- | --- | --- |
| • Anonymous data | • Pseudonymised normal personal data | • Pseudonymised special category personal data<br>• Additional safety measures<br>  • encryption<br>  • ethical approval | • Non-pseudonymised special category data<br>• Additional safety measures of risk level 2<br>• Extra layer of organisational and technical security<br>  • restricted access rights (role based)<br>  • detailed audit trails<br>  • strong authentication |

# How to pseudonymise?

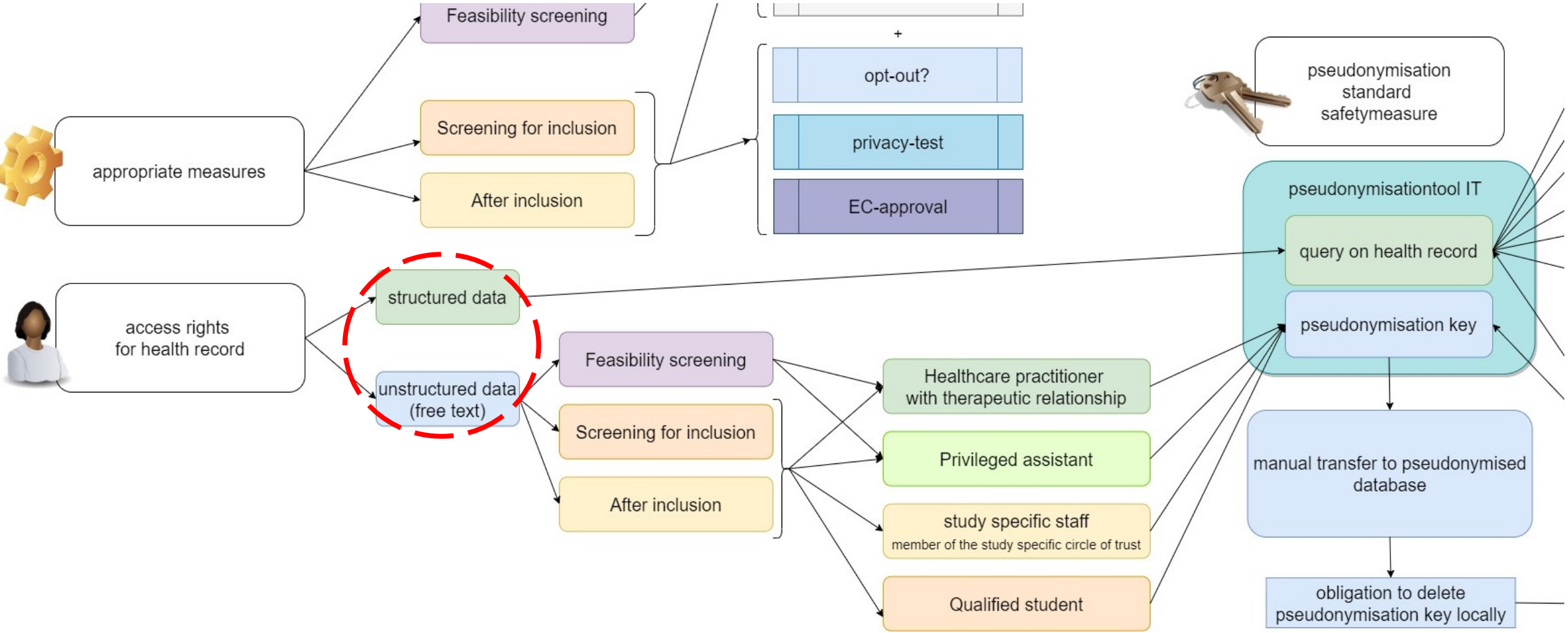
1. Replace the patient ID with a study-specific ID
   - Avoid the use of cross-study identifier to reduce linkability

2. Remove / replace / generalise,… other identifiers
   - 18 identifiers HIPAA for de-identification as rule of thumb
   - But only to achieve pseudonymised, not anonymised, dataset

# 18 HIPAA identifiers

1. Name
2. Address (all geographic subdivisions smaller than state, including street address, city county, and zip code)
3. All elements (except years) of dates related to an individual (including birthdate, admission date, discharge date, date of death, and exact age if over 89)
4. Telephone numbers
5. Fax number
6. Email address
7. Serial number or unique identifier of (medical) device
8. Social Security Number
9. Medical record number
10. Health plan beneficiary number
11. Account number
12. Certificate or license number
13. Any vehicle or other device serial number
14. Web URL
15. Internet Protocol (IP) Address
16. Finger or voice print
17. Photographic image - Photographic images are not limited to images of the face
18. Any other characteristic that could uniquely identify the individual

# Structured versus non-structured data



# Conclusion

- Article 89§1 introduces a three-level cascade:
  anonymization → pseudonymisation → non-pseudonymised data

- Pseudonymisation is an essential measure to protect research participants
  - Reduction of the risk for the research participant to be identified
    - Encoding is just one step
    - Other “anonymization” techniques have to be applied
  - No obstruction to empowerment of the research participant

- Because of the reduction of risk for re-identification pseudonymisation affects access management, storage and information security policies for scientific research

dr. Griet Verhenneman
DPO University Hospitals Leuven
Lecturer European Privacy and Data Protection Law KU Leuven
Affiliated Researcher CiTiP – KU Leuven

griet.verhenneman@uzleuven