Applies to question C4: IDENTIFIABILITY in RESULTS: CAN PEOPLE BE IDENTIFIED OR RE-IDENTIFIED IN THE ANALYSIS DATA AND/OR RESULTS? We first explain the basic aspects of anonymity and pseudonymity and then explain what you need to answer when opting for YES or NO to this question.

NON-ANONYMOUS VS. FULLY ANONYMOUS RESEARCH

Fully Anonymous Research = "no one is recognizable by anyone - with the use of reasonable resources"
- Participants are never identifiable at any stage, including recruitment. For example, no emails, names, or other identifiable details are collected or used.
- If a participant can be recognized by even one researcher, the study is non-anonymous.
- Street surveys or quick public polls can sometimes qualify as anonymous if the researcher doesn't know participants and collects no identifiable data.
- 'With the use of reasonable resources': all those who can trace individuals in data files with special equipment or actions are not included here.

Non-anonymous Research = Participants are identifiable at some point, such as during recruitment (e.g., via email) or data collection (e.g., recording names, audio, or video).
- Some examples of non-anonymous research:
    o Researchers come into contact with participants.
    o Recordings of faces or voices are collected or made.
    o Contact details are used or collected.
    o (Other) identity data are requested.
    o Data or combination of data can lead to the recognition of a participant.

A study cannot be both anonymous and non-anonymous. Use the correct term in your application.

NON-ANONYMOUS, ANONYMOUS VS. PSEUDONYMOUS DATA

Non-anonymous data = Participants are directly or indirectly identifiable in the dataset.
- *Direct identification*: Data include names, contact details, audio or visual material and the like.
- *Indirect identification*: based on the combination of some data individuals can be traced. For example, the name of a company and the position of an employee, or the age of an outlier in a population, etc. It is *your responsibility* to estimate whether participants can be recognizable based on the data you collect.
- When in doubt, consider the data to be non-anonymous.

Anonymous data = No one can identify or re-identify participants.
- You have never collected data with which one can directly or indirectly recognize someone, OR
- You *anonymize your data*: You have collected data with which participants can be recognized, but you permanently delete this data. (For example, destroying audio recordings after transcription where participants are no longer recognizable in the transcripts).

Pseudonymous data = Participants are re-identifiable using a coding system.
In this case there is:

1. A database in which participants are recognizable,
2. A data file in which participants are made unrecognizable with a code, and
3. A key file that can link the code from file 2 to the data from file 1.

Data cannot be both anonymous and non-anonymous or pseudonymous at the same time. Use the correct terms in your application.

Non-anonymous data can be made pseudonymous or anonymous.

Anonymization = Data is permanently modified so individuals can never be identified again, directly or indirectly.
- For example, identifiers like names or recordings are deleted after anonymized transcripts are created.
- Re-identification is not possible.

Pseudonymization = Data is modified by replacing identifiers (e.g., names) with codes, but individuals can still be re-identified using a separate key file that links the codes to the original data.
- Re-identification is still possible.

Data cannot be both anonymous and non-anonymous or pseudonymous at the same time. Use the correct terms in your application.

Also look at the visual flowcharts at the end of this document.

THUS:
- **Prefer anonymous data** collection if this is possible,
- **Prefer the anonymization** of data to pseudonymization,
- **Use the terms correctly** and consistently throughout your application,
- **Inform your participants correctly**:
    - Is your data collection anonymous or not?
    - If not: will you anonymize, pseudonymize or keep data non-anonymous? Explain why you choose which option, and from which point data may or may not become anonymous/pseudonymous. ALWAYS USE THE CORRECT TERMS! If not participants will be confused and you come across as unprofessional.

YES TO QUESTION 4

YES to question 4 = people can be identified or re-identified in the analysis data and/or results. In this case, address the following:

1. WHY are the data not anonymous in the analyses/results?
    - Clearly justify why participants are or remain identifiable in the analysis data or results.
    - Avoid vague justifications like "necessary for research" or "required for open science." Be precise about the value and relevance of identifiability to your research outcomes.

- Examples:
  - o Identifiability is essential for understanding individual case studies (e.g., participant narratives in qualitative research).
  - o The study requires linking responses to specific individuals across time or contexts (e.g., longitudinal studies).
- Be aware that Horizon Europe's open science policy does not require that all data should be accessible for all research purposes. Especially for non-anonymous data, they too emphasize the need to protect participants' privacy. See https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf
- **Make it specific and explicit that recognizability is necessary** ensures that participants are fully informed of its purpose and potential implications, aligning with ethical and GDPR standards.

2. WHICH data/study results can be linked to individuals?
- Specify the data or results that may reveal participants' identities, or that could be linked to their identity.
- This ensures transparency with participants and allows the ethics committee to assess the potential risks of recognizability in your study data/results.
- We are especially cautious with data/results that may be perceived as taboo or sensitive to the participants.

3. HOW long do people remain recognizable in the data?
- State the exact period and justify why remain recognizable for that duration.
- GDPR requires identity data to be stored only as long as necessary. Keeping recordings indefinitely increases risks to participant privacy.

4. WHO can consult data in which people are recognizable?
- List everyone with access (e.g., researchers, assistants) and their roles.
- Participants have the right to know who can view their identity data; you must also mention this information in the informed consent.
- Limiting access reduces risks of misuse or breaches: restrict access to non-anonymous data.

5. Do all participants give ACTIVE CONSENT for the recognizability in analysis data/results? Please arrange this in the informed consent form, this requires separate explicit consent.
- In the INFORMED CONSENT:
  - o Clearly explain that participants will be recognizable in data/results. Also explain who can access their identity information and for how long. It is important to explain why this is necessary for the research and what may be the consequences for the participants.
  - o Ensure that participants give ACTIVE CONSENT for the recognizability in analysis data/results. PLEASE NOTE: Only use relevant and CUSTOMIZED CONSENT OPTIONS from our website's informed consent forms. Unclear, missing or redundant information in the informed consent forms will result in a PRELIMINARY NEGATIVE ADVICE from EASHW.
- Use the INFORMED CONSENT TEMPLATE FOR NON-ANONYMOUS RESEARCH.
- This template includes only the minimum required information as per EASHW and GDPR standards. While you must adapt the language to suit your target population, the content itself cannot be changed.

6. Do all researchers involved (including external persons and (job) students) sign a CONFIDENTIALITY DECLARATION? Attach these declarations.
- Make sure that everyone with access to non-anonymous data signs a CONFIDENTIALITY AGREEMENT, and ensure these documents are attached to your application.

NO TO QUESTION 4

NO to question 4 = people can NOT be identified or re-identified in the analysis data and/or results. In this case, address the following:

Specify whether your study is:
1. A PRIORI FULLY ANONYMOUS:
- No identifiable information is ever used or collected during the study.
- Explain how participants remain unidentifiable at every stage of the research.
- For example:
  - Anonymous surveys conducted without collecting names, email addresses, or demographic data that could indirectly identify individuals.
  - Observations in public spaces where individuals cannot be distinguished or identified.

OR

2. INITIALLY NON-ANONYMOUS BUT LATER ANONYMIZED:
- Identifiable data is collected during the study but anonymized before data analysis begins.
- In this case, explain:
  - What data was initially identifiable: (e.g., interview recordings, survey responses linked to email addresses).
  - How and when anonymization occurs: (e.g., deleting recordings, removing names or identifiers).
  - Why anonymization is irreversible: Ensure no key file or mechanism exists to re-link anonymized data to individuals.

This is a visual representation based on a fictional example:

Original non-anonymous dataset

Full dataset
Name
Email
Employer + position
Gender
Nationality
Average number of steps per day
Average number of hours of sleep
Average number of publications per year

Data with which participants are recognizable:
Name
Email
Employer + position

Data with which participants are not recognizable
Gender
Nationality
Average number of steps per day
Average number of hours of slack
Average number of publications per

Pseudonymization =

Data with which participants are not recognizable
CODE
Gender
Nationality
Average number of steps per day
Average number of hours of slack
Average number of publications per

➕

Data with which participants are recognizable (KEY file):
CODE
Name
Email
Employer + position

🟰

Full dataset
Name
Email
Employer + position
Gender
Nationality
Average number of steps per day
Average number of hours of slack
Average number of publications per

Saved to protected file on OneDrive
May be shared

KEY FILE
Saved to SEPARATE protected file on OneDrive
May NOT be shared

Re-identification is possible

Anonymization =

Data with which participants are not recognizable
CODE
Gender
Nationality
Average number of steps per day
Average number of hours of slack
Average number of publications per

Data with which participants are recognizable:

Full dataset

Gender
Nationality
Average number of steps per day
Average number of hours of slack
Average number of publications per year

Saved to protected file on OneDrive
May be shared

PERMANENTLY DELETED
(no copy kept anywhere)

You only keep an anonymous dataset