

NEAR - Rules and Guidelines for Data Documentation, Harmonization, Access and Storage

Introduction	2
1. NEAR	2
2. Aims of the document and target readers	2
3. Terms and definitions.....	3
Data Documentation.....	3
Data Harmonization.....	4
Data Access (Terms of use).....	4
1. Scope and Subject matter	4
2. Terms of Use and eligibility criteria	4
3. Intellectual property rights and User licenses.....	5
3.1 Ownership of the NEAR- Data	5
3.2 NEAR User licenses	5
4. Information duties of the applicant/user	6
5. Notification, acknowledgement, and authorship.....	7
5.1 Authorship.....	7
5.2 Acknowledgement requirements.....	7
5.3 Notification requirement.....	8
6. Data protection and privacy rules	8
6.1 Anonymity of participants.....	8
6.2 Applicable law	8
7. Confidentiality rules	8
8. Ethics	8
9. Disclaimer	9
10. Consequences of breach of the NEAR Conditions of Use	9
11. Update of the NEAR Conditions of Use	9
Data Storage.....	10

INTRODUCTION

1. NEAR

This document articulates the minimum conditions for access to datasets included in whole or in part in NEAR.

NEAR is a unique research project that was founded in 2017 by the Swedish Research Council (Vetenskapsrådet, VR) with the overall aim to build and run a *National E-infrastructure for Aging Research* (“NEAR”). NEAR is a collaboration between seven universities including Karolinska Institutet, University of Gothenburg, Lund University, Umeå University, Jönköping University, Uppsala University, and Blekinge Institute of Technology. NEAR is a multidisciplinary resource that comprises 15 major longitudinal population-based studies on aging and health in Sweden (“local databases”) and includes 180.000 individuals aged 50 years or over who have been assessed by direct interview and clinical examination, and followed between 7 and 52 years (2020).

NEAR includes the following local databases:

- Betula Project, Umeå University
- The Swedish National study on Aging and Care – Nordanstig (SNAC-N), Karolinska Institutet
- The Cohort of Swedish Men (COSM), Uppsala University
- The Swedish Mammography Cohort (SMC), Uppsala University
- Kungsholmen Project (KP), Karolinska Institutet
- The Swedish National study on Aging and Care – Kungsholmen (SNAC-K), Karolinska Institutet
- H70 Birth Cohort Studies (H70, H85, H95+), University of Gothenburg
- The Swedish National study on Aging and Care – Blekinge (SNAC-B), Blekinge Institute of Technology
- The Swedish National study on Aging and Care – Skåne (SNAC-S), Lund University
- Gender Differences in Health Behaviour and Health among Elderly (GENDER), Jönköping University
- Origins of Variance in the Old-Old (OCTO-Twin), University of Gothenburg
- Screening across the Lifespan of Twins (SALT), Karolinska Institutet
- Swedish Adoption/Twin Study of Aging (SATSA), Karolinska Institutet
- Survey of Health, Ageing and Retirement in Europe-Sweden (SHARE-Sweden), Umeå University
- Swedish Panel Study of Living Conditions of the Oldest Old (SWEOLD), Karolinska Institutet

KI is the coordinator of NEAR.

The NEAR participants have agreed upon the terms of the collaboration in a consortium agreement.

2. Aims of the document and target readers

The aim of this document is to provide:

- a) **Internal regulations** for the Principal Investigators of the study cohorts included in NEAR; and
- b) **Terms of Use** for potential applicants (“Users”) of NEAR.

3. Terms and definitions

- *NEAR central*: the meta-database including documentation and harmonization derived from the participating studies in NEAR. It is driven by the operational group and it is located at Karolinska Institutet.
- *Local databases*: The databases derived from the individual study cohorts that are included in NEAR and geographically located all around Sweden at their respective universities (as previously listed).
- *Principal Investigator (PI)*: researcher responsible of the cohort study and related local database, part also of the steering group of NEAR.
- *Applicant*: any researcher who applies for NEAR-data.
- *User*: Any applicant granted to use NEAR-data for analytical purposes.
- *Project*: A well-defined proposal, which results in limited number of outputs specified in the application (e.g. reports, journal papers).
- *Metadata*: variables description in the e-catalogue.
- *Processed variables*: All the raw data collected through questionnaires, interviews, and clinical examinations, but cleaned and transformed into a format that can be analyzed and/or visualized.
- *Derived variables*: Variables derived from the processed variables and include the diseases diagnosis (e.g. dementia), imaging (e.g. MRI), scales (e.g. personality profiles), scores (e.g. physical activity levels), health- related indexes, and latent variables.
- *NEAR-data*: *harmonized data* processed and/or derived variables from two or more of the study cohorts included in NEAR, which have been created after an accepted NEAR application.
- *NEAR metadata*: all information regarding study descriptions and data dictionaries in an electronic format for each cohort and every data collection event from each local database.

DATA DOCUMENTATION

NEAR central will document NEAR metadata.

The documentation will cover all relevant variables (Process and Derived variables) that are already stored in the local databases and will be stored in the future as long as the local database is part of NEAR. Specifically:

The documentation of all NEAR metadata listed and published by NEAR is available free of restriction.

DATA HARMONIZATION

Data will be harmonized by NEAR central after approval of local database-PI to use NEAR-data. The harmonization will be performed in collaboration between NEAR central and the relevant PI. Specific technical solutions will be applied to meet all the needs of the local databases. The harmonization will cover only the processed data. If not yet done, the cleaning will be carried out at the local databases.

Each User of an approved application will be involved in the harmonization process.

The harmonization will proceed in accordance with the Maelstrom guidelines for harmonization, and will be supervised by Maelstrom. The procedure includes six steps:

Step 0. Define the research question(s), objectives, and protocol;

Step 1. Assemble information and select aims

Step 2. Define variables and evaluate harmonization potential

Step 3. Process data

Step 4. Estimate quality of the harmonized dataset(s) generated

Step 5. Disseminate and preserve the final harmonization products

For a more detailed description of the Maelstrom Guidelines for harmonization, see:

<https://www.maelstrom-research.org/about-harmonization/maelstrom-guidelines> or

<https://academic.oup.com/ije/article/46/1/103/2617181>.

DATA ACCESS (TERMS OF USE)

1. Scope and Subject matter

The present Terms of Use apply to the "NEAR-data". All Users are required to agree to the Terms of use of the NEAR-data.

NEAR data will be made available under certain conditions according to the type of variables requested and the type of User (cf. 2, below).

All the *Processed variables* and *Derived variables* are included in the NEAR metadata.

A Fee for NEAR-data and/or samples will be requested depending on the kind of User (academic or non-academic), the quantity and the quality of the data requested. Fees will depend on services provided (harmonization, analyses, coordination) and be paid to NEAR central (KI), the fee will be established by the NEAR Steering Board. Some databases are included in other infrastructures, the fee to these databases will be paid directly to the database.

2. Terms of Use and eligibility criteria

All Users of NEAR-data will have to comply with the following eligibility criteria and basic Terms of Use. The NEAR-data may be used for both scientific and non-scientific purposes e.g. governmental authorities, non-profit organizations or companies (pharma). However, the local NEAR (database) have the final decision on who will get to use their data. Reasons for not allowing data usage needs to

be strongly motivated. Different requirements are applied according to the User's membership in either an Academic (i.e. university or research institute) or Non-Academic institution (health care professionals; industrial sector, government agencies, and national/international organizations). Restrictions may apply to non-Academic institutions for the use of NEAR-data from specific cohorts.

All potential Users requesting access to NEAR-data need to submit an application by filling-in and entering a specific form available at the NEAR website (<https://www.near-aging.se/>) and, when applicable, at the Local Databases' homepage.

The application has to contain relevant details about the study design of the project for which the data shall be used; the ethical permission to perform the project or proof of the submission of the ethical application to the Swedish Ethical Review Authority; and the curriculum vitae of the applicants.

Applications for non-scientific purposes will not need an ethical approval, for definition of scientific purposes see Swedish law (2003:460).

Applicants who succeed in providing the required information are eligible to register as NEAR Users. Failure to comply with the eligibility criteria and basic terms of use or to provide the relevant information will result in a rejection of the application.

The NEAR Scientific Coordinator will regularly follow up the project once the application has been approved throughout the duration of the project. The PIs of the involved local databases may also request regular updates from the NEAR Scientific Coordinator. The PIs of the involved local databases have the right to withdraw access to their data from a project (with or without scientific purpose) after such access has been approved. The denial takes place in case where the legitimate interests of the local database would be severely affected.

3. Intellectual property rights and User licenses

3.1 Ownership of the NEAR- Data

As laid down in the NEAR Consortium Agreement (accessible at NEAR webpage), each local database continues to be the owner of a newly generated variables as a result of a NEAR project, and intellectual property rights related thereof.

The Users of the NEAR infrastructure shall not be granted any ownership or intellectual property rights to the local databases or to the NEAR-data.

3.2 NEAR User licenses

The NEAR User will obtain the right to use the NEAR-data ("user license") upon approval of an application, only for the granted project and for the duration of the project. Such user license will end automatically when the project or the related publication is completed.

Users employed at or affiliated to a University or related research institutions will receive a data-file with the requested variables, built in accordance with current data and privacy protection regulations (see paragraph nr 6). Specific technical solutions will be implemented in order to meet any requirements of the local databases. All the NEAR-data, including new variables generated by the User

and script to run the analyses reported in scientific journals, have to be returned to NEAR central upon completion or termination of the project, in order to be archived according to NEAR rules (see paragraph “Data storage”). Restriction to the generation of new variables might apply for specific cohorts (e.g. SHARE-Sweden).

The NEAR User will only have the right to archive the script to run the analyses. All the NEAR-data including the generated data shall be destroyed by the User upon termination of the project. A certificate of destruction will be provided to NEAR/KI.

Users outside Universities or related research institutions will receive the requested information as analyzed results. The analyses will be carried out by NEAR statisticians under the supervision of the NEAR Scientific Coordinator.

NEAR Users shall only use the NEAR-data or the analyzed results for the duration of the approved project. The duration of a project may be extended upon formal request, in duly justified cases.

NEAR Users may not submit a new proposal for a subsequent project, until the analyses of the current project have been completed and the first draft of the report is made available. The draft report shall be attached to the new application.

4. Information duties of the applicant/user

As part of the registration as a NEAR User certain personal and professional details have to be provided by the applicant. By signing the Data Transfer Agreement (DTA) the applicant warrants that any information submitted by him/her is accurate and truthful and that he/she will keep this information accurate and up to date after access has been granted. In the event that any of the details provided to NEAR change (such as email address and scientific affiliation), the User shall promptly inform NEAR. All changes will be treated separately and approval is subject to approval of the Term of use.

If the User application is approved any further use of the data for other purposes other than the NEAR PROJECT, must be approved separately, upon submission of a new application.

Should an applicant or a User provide false, out-of-date or incomplete information or should there be reasonable doubt regarding the truthfulness of the information provided, NEAR reserves the right to a) reject the application,

b) suspend or block access to the data or

c) withdraw the right of use the NEAR data

with immediate effect.

Outputs of projects using NEAR data should be published in on-line open access journals, or in journals that allow open access, pursuant to the rules of the Swedish Research Council.

5. Notification, acknowledgement, and authorship

Any planned publication using NEAR-data (printed as well as online), shall comply with the following notification and citation obligations:

5.1 Authorship

Authorship on publications of NEAR-projects will be based on academic standards and custom. In accordance with normal academic practice, all investigators and contributors to a publication will be acknowledged, always in compliance with recognized standards concerning publication and authorship, including the most recent “Recommendations for the Conduct, Reporting, Editing and Publications of Scholarly Work in Medical Journals” developed by the International Committee of Medical Journal Editors (ICMJE).

In particular, authorship on publications based on NEAR data published in scientific journals will be based on the following four criteria:

- Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
- Drafting the work or revising it critically for important intellectual content; AND
- Final approval of the version to be published; AND
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

The User of the approved project is responsible for identifying who meets these criteria and co-authorship will be agreed upon preparation of the project application. However, modifications are possible during the implementation of the project subject to approval by the NEAR Scientific Coordinator. The PI of the local databases involved in the project will be asked to participate in the project either directly or through another collaborator/s. The PI of the local database and/or the collaborator will become co-authors if the above criteria are fulfilled.

5.2 Acknowledgement requirements

The NEAR data set(s) and their main funding shall be acknowledged in all publications using NEAR-data. Funding of a local database included in the approved project that should be acknowledged will be provided by the PI of each local database.

The following NEAR acknowledgement standard text must be included in all publications of NEAR-approved projects:

This Study uses data from the national infrastructure NEAR, supported by the Swedish Research Council (Dn 2017-00639 and 2021-00178). Specifically, the following NEAR databases have been included: DATABASE 1(Dn xxxx) , 2, ect....

5.3 Notification requirement

As soon as there is an official reference (including working papers, theses, etc.) based on NEAR data, it needs to be reported to the central NEAR Scientific Communicator. This information will be published on the NEAR website (online bibliography database).

6. Data protection and privacy rules

6.1 Anonymity of participants

NEAR-data are released in the form of "scientific-use files" protected by factual anonymity (as defined by EU:s General Data Protection Regulation (GDPR) <https://gdpr.eu/>). Such data may be used for both scientific and non-scientific purposes. NEAR data shall never contain identifying information about the NEAR participants.

The User shall not use the NEAR-data to identify the data subjects nor to contact them under any circumstances. Recipient shall refrain from analyzing and/or using NEAR-data in a way that has the potential to (i) lead to the re-identification of any data subject, (ii) compromise the anonymity of any data subject in any way, or (iii) result in use of NEAR-data outside the strict scope of any data subject's consent or the scope of the approval of the application.

6.2 Applicable law

The use of NEAR-data is subject to European Union and national data protection laws. NEAR Users shall sign appropriate data protection agreements depending on their geographical location with the University/ies hosting the databases.

7. Confidentiality rules

Users are required to take appropriate technical and organizational measures to protect the copies of the NEAR-data against unauthorized access.

Users are not allowed to make copies of the data available to others and/or enable any third-party access to the NEAR-data. Access to the NEAR-data is only granted on an individual/team basis.

Detailed information on confidentiality rules will be provided in the Data Transfer Agreement.

8. Ethics

Ethical approval is needed for any research that involves human participants; their tissue and /or data to ensure that the dignity, rights, safety and well-being of all participants are the primary consideration of the research project.

The data collection procedures of the local databases included in NEAR are subject to continuous ethical review and approval. NEAR's activities related to human subjects-research follow international research ethical principles such as the 'Declaration of Helsinki'.

All NEAR Users are expected to be familiar with the fundamental principles of research ethics (see guidelines mentioned above) and to take them into account in an appropriate manner when conducting research using NEAR-data.

NEAR Users shall ensure that its activities under the project conducted with NEAR-data fully comply with the fundamental principles of research ethics, such as, any World Medical Association policy included in the Handbook of WMA Policy, for instance The Helsinki Declaration on Ethical Principles for Medical Research involving Human Subjects, International Code of Medical Ethics, Declaration on Ethical Considerations regarding Health Databases.

NEAR Users shall secure all necessary approvals from the relevant research ethics committees in their geographical location before undertaking any part of the project conducted using NEAR-data.

NEAR may require proof that ethical approvals are obtained and may perform inspections of the User's observance of this condition. NEAR may terminate the cooperation and any future agreement with immediate effect due to the User's breach thereof.

Handling data for non-scientific purpose does not require ethical permits. Definition of scientific purpose can be found in the Swedish law 2003:460.

9. Disclaimer

NEAR bears no responsibility for the use of the NEAR-data, or for interpretations or inferences based on this use. The responsibility for Users' research, projects and publications using NEAR-data and, in particular with regard to the research results, lies solely with the User.

No warranty or representation is made of any kind is made, given or implied as to the sufficiency or fitness for purpose nor as to the infringement of any proprietary rights of third parties in respect of any data supplied by NEAR to the User. Therefore, the User shall in all cases be entirely and solely liable for the use to which it puts the NEAR-data.

Users shall be solely liable for any loss, damage or injury to third parties resulting from the use of NEAR-data.

10. Consequences of breach of the NEAR Conditions of Use

In the event of breach of the Terms of Use, NEAR has the right to withdraw the right of use of the NEAR-data with immediate effect from the user at any time and request him/her to delete all copies of the data immediately.

In case of an intentional or serious breach of the data protection and privacy rules, or attempts thereof, NEAR reserves the right to make public the data protection violation including the identity of the user.

11. Update of the NEAR Conditions of Use

The NEAR Terms of Use may be subject to change due to e.g. legislative amendments, updates of NEAR policies or changes in the funding conditions.

In such cases, NEAR may, from time to time, vary, amend or delete any provision of these Terms of Use without further notice. Users of ongoing NEAR approved projects will be informed via email

about the updated Terms of Use. Unless otherwise stated in the email, the updated Terms of Use shall apply as from the first calendar day from date the email is sent.

DATA STORAGE

Original, harmonized, processed, and Derived variables remain stored at the local database under the responsibility of the University where the dataset is located, and the respective cohort PI. Different agreement can be made separately between NEAR and the individual local databases on specific variables.

No research data files with individual data will be archived at NEAR central. Analytical syntax files and syntax files for creating data to carry out the approved NEAR projects will be stored at NEAR central for a period of 10 years according to archival laws and research documentation regulations. These files will be safely stored and only accessible upon prior written approval by the PI's involved.