Plan Overview

A Data Management Plan created using DeiC DMP

Title: NORA: Navigating Organic and Refinement's impact on human and planetary health

Creator: Sophia Leiss

Principal Investigator: Sophia Leiss

Data Manager: Sophia Leiss

Project Administrator: Senior Scientist Cecilie Kyrø

Affiliation: Københavns Universitet / University of Copenhagen

Funder: European Commission

Template: UCPH Data Management Plan template

Project abstract:

NORA is a multi-partner research project that brings together epidemiology, clinical trials, omics, and environmental life-cycle assessment to study how processing and production methods shape health and sustainability. The current PhD will contribute the human-health sub-studies. Environmental impact analyses are conducted by other NORA collaborators.

Plant-based diets can benefit health and the environment, but two risks must be addressed: ultra-processed plant products may carry adverse health effects, and higher intake of conventionally grown plants can increase exposure to pesticide residues. NORA will evaluate these dual dimensions by jointly considering the degree of processing and organic versus conventional production.

The objective is to quantify how processing of plant-based products and organic production relate to cardiometabolic diseases, and to identify oral-gut microbial and metabolic pathways that explain these associations.

First, a large register-linked cohort study based on the Diet, Cancer and Health – Next Generations cohort will assess associations between consumption patterns and cardiometabolic outcomes, using a standardized NOVA protocol for processing degree and recorded organic use. Second, a randomized crossover trial will test plant-based diets differing in processing level, measuring short-term health effects on cardiometabolic risk factors, and oral–gut microbiome and metabolomic profiles. Insights from the trial will be used to derive biomarker signatures that may be applied in future cohort work to strengthen inference. Environmental impact estimates will be produced within the broader NORA project.

ID: 8197

Start date: 01-05-2026

End date: 30-04-2029

Last modified: 18-05-2026

Grant number / **URL:** 101179234

NORA: Navigating Organic and Refinement's impact on human and planetary health

0. Project information

0.1 Workzone Case Number

0.2 Project Members (including collaborators)

The project will mainly be carried out at the DCI with Cecilie Kyrø and collegues. During the PhD, there will be a research stay in Sweden in 2027 with the Swedish collaborator Rickard Landberg (Chalmers University in Gothenburg) to conduct a clinical trial.
Main supervisor, Prof. Torben Hansen (KU NNF Basic Metabolic Research) and co-supervisor Prof. Susanne Pedersen (DTU Bioengineering) will give conceptual guidance and input.

0.3 Other Project Documentation

The overall NORA project is funded by the Danish Cancer Institute's (DCI) approved application to Plantefonden. The PhD project is partially funded by the PhD Fellow Grant Agreement HE-MSCA-COFUND [No 101179234] - Marie Skłodowska-Curie Ac-
tion Networks (HE MSCA-COFUND) - Acronym: [INTERACT].

1. Data description

1.1 Describe what material and data will be collected, observed, generated, created or reused in the project. For the different types of research data, address their:

- **Origin** / **Source**
- **Estimated size** / **Volume**
- **Expected format(s)**

Register study:
We will use the data from the Diet, cancer, and Health - Next Generation cohort which will be requested from the Danish Statistic Institute, and is part of a large european, epidemiological data collection. The data has already been collected between 2015-2019 and includes about 40.000 participants. It included data collection in participant centers and filling out a food frequency questionnaire as well as a lifestyle questionnaire. When we request certain variables for our NORA/ PhD project, our data managers will clean the data first before it is transferred to the PhD student for analysis.
Different types of variables are expected: continous, discrete, categorical, binary, etc.
Expected formats: CSV/TSV, Stata .dta, R .rds/.parquet; documentation in PDF/Markdown; code in .R/.do/.py.

Trial data in Gothenburg in 2027:
The PhD student will not collect the data themselves, but will be responsible for project management of the trial as well as analysis, which includes metabolomics/proteomics data derived from fecal, saliva, and blood samples:
Clinic measurements (anthropometry, BP), fasting/postprandial biomarkers, questionnaires (adherence, symptoms), and diet logs.
A sample size of 60 participants is currently expected. This trial will be a feeding study with a cross-over trial design for about 6 weeks. These data is typically treated as a multivariate continous data but requires preprocessing (conducted by Swedish team) before standard methods apply.
Expected formats: CSV/TSV, REDCap/eCRF exports (CSV/XML), Stata .dta, R .rds; SOPs in PDF.

1.2 Describe any material and data that contain sensitive information, in particular:

- **Personal data**
- **Human biological material, including biobanks**
- **Classified information**
- **Confidential (business) information**
- **Any other material or data that must be protected to safeguard the security of individuals, organisms, communities, organizations, etc.**

Personal data The cohort (DCH–NG) and the randomized cross-over trial both contain person-level data that are personal and health-related (special category under GDPR). Variables include age, sex, education, income bands, cohabitation, lifestyle, and clinical measurements. All analysis datasets are pseudonymized; key files are stored separately with restricted access.
Human biological material (biobanks) The trial collects human biological material (blood, saliva/oral swabs, stool). Samples are coded, stored at –80°C under SOPs, and linked to participants only via a secure key file. Any residual DCH–NG biobank samples used for ancillary assays are handled under existing cohort governance and approvals.
Facility security: sample storage locations and detailed IT architecture are not disclosed publicly.
Third-party data: any register linkages are processed under data-processing agreements; sharing is limited to what approvals permit.

2. Rights to research data

2.1 Address whether there are any access restrictions to material and data during the project. If so, describe who can have access to the material and data during the project, under which conditions, and in what timeframe.

- Legal/ethical framework: All person-level data are processed under GDPR, Danish Data Protection law, and cohort/trial approvals. Access follows least-privilege, role-based control.
- DCH–NG cohort data: Primary FFQ/lifestyle data and consented register linkages are governed by DCH–NG. Access is restricted to named investigators on approved protocols, via secure analysis environments. No raw person-level data leave these environments. External collaborators receive only approved extracts or aggregate outputs.
- Trial clinical data: Access is limited to the trial PI, data manager, and named analysts listed on the ethics and data-processing approvals. Data reside in a secure eCRF/REDcap instance

and institutional servers; exports are pseudonymized. Monitoring/auditing logs are enabled. Unblinding (if any) follows the statistical analysis plan.

- Biobank samples and omics raw data: Samples are coded; release requires PI approval and documented laboratory SOPs. Raw omics files (FASTQ, mzML) are stored on restricted servers; access is granted to named analysts under a data-use agreement. Only processed, de-identified feature tables are shared more broadly within the team.
- Third-party/contractor access: Sequencing or MS vendors access only coded samples and minimal metadata under service agreements that prohibit secondary use.
- Timeframe: Access is granted for the project duration and reviewed at major milestones (e.g., first data freeze, manuscript submission). Departing team members lose access; archives remain under PI/institutional control.

2.2 Describe any material and data in your project that are subject to intellectual property rights.

- Copyright (third-party): DCH–NG questionnaires, codebooks, and person-level cohort/register data are copyrighted and governed by cohort institutions. We use them under approved protocols; redistribution is not permitted. Journal articles will be under publisher copyright (we will prefer open access where possible).
- Copyright (our outputs): NOVA coding protocol text, derived exposure dictionaries, and analysis code created in this project are our IP. Plan: release after publication under open licenses (CC-BY for documentation; MIT/BSD for code), excluding any confidential/person-level data.
- Patents/designs: No patentable inventions or registered designs are anticipated.

2.3 List any agreements or contracts set up in your project that contain provisions on rights to material and data, such as research collaboration agreements, non-disclosure agreements, material transfer agreements or license agreements.

- DCH–NG Data Access Agreement and Cohort Governance approvals
  Grants right to use person-level FFQ/lifestyle data and consented register linkages within approved protocols; prohibits redistribution; defines publication and acknowledgment rules.
- Register linkage Data Processing Agreements (DPAs)
  GDPR-compliant DPAs with relevant Danish authorities for linkage and secure processing of sociodemographic variables; specify lawful basis, roles, retention, and access controls.
- Ethics approvals and Participant Information/Consent Forms (ICFs) for the trial
  Define permissible uses of clinical data and biospecimens (blood, saliva, stool), storage, secondary use, and withdrawal rights; include rights to publish aggregated results.
- Biobank/Material Transfer Agreements (MTAs)
  Cover transfer of coded samples to certified laboratories (microbiome sequencing, metabolomics); prohibit secondary use; require destruction/return after analysis.
- Software and database licenses
  Institutional licenses for analysis platforms; open-source licenses for third-party libraries; terms for eCRF use.
- Research collaboration agreements (Danish partners, WPs)
  Define roles, foreground/background IP, authorship, data sharing of aggregates/derived variables, and embargo periods.
- Non-disclosure agreements (as needed)
  With partners for pricing, configurations, or unpublished methods.
- Publication and data-sharing policy (project-level)
  Internal policy aligning with cohort governance and GDPR, covering preprints, code release, and conditions for sharing de-identified derived datasets.

Note: No agreements permit public release of person-level cohort or trial data; only de-identified aggregates/derived outputs and code will be shared per approvals.

2.4 List any legislation, policies, guidelines or requirements that govern research data management in your project.

- GDPR and Danish Data Protection Law
  - EU General Data Protection Regulation (GDPR) and the Danish Data Protection Act govern all person-level cohort and trial data.
  - Danish Data Protection Agency guidance and institutional Data Processing Agreements (DPAs) for register linkages.
- National research ethics and biobank governance (Denmark)
  - Danish Act on Research Ethics Review of Health Research Projects and related executive orders.
  - Declaration of Helsinki for ethical conduct.
  - Rules for collection, coding, storage, and secondary use of human biological material (institutional/biobank SOPs; approvals from relevant biobank governance).
- Institutional policies (University of Copenhagen)
  - UCPH Policy for Research Data Management (2022), including data classification, storage, retention, backup, and sharing.
  - UCPH/host-institution IT security policies (access control, encryption, secure transfer).
- Cohort and data-provider regulations
  - DCH–Next Generations cohort governance and data-use agreements (no redistribution of person-level data).
  - Register authorities' terms for linkage, processing, retention, and disclosure control.
- Clinical research standards (trial)
  - SPIRIT (protocol) and CONSORT (reporting) guidelines, including crossover-trial extensions where applicable.
  - Good Clinical Practice principles (monitoring, documentation, SAE reporting) as required for non-drug dietary trials by local ethics/Institutional policies.
- Laboratory and analytical standards
  - SOPs for collection/processing/storage of blood, saliva/oral swabs, and stool; temperature monitoring and chain-of-custody.
  - Quality standards followed by service labs (e.g., ISO 15189 for clinical labs, validated methods for LC-MS/sequencing).
- Funder and partner requirements
  - Plantefonden grant conditions (Open Science, DMP updates, acknowledgments).
  - Collaboration agreements defining IP, authorship, data sharing of aggregates/derived variables.
- Publisher and disciplinary requirements
  - Journal data/code availability policies; anonymization and disclosure-control for outputs.
  - STROBE for observational analyses; nutrition-specific checklists where relevant.

2.5 Describe whether, when and how research data may be used for other purposes (e.g. other research projects) and what arrangements will be made if a project member leaves the project and/or UCPH before the end of the project.

- Reuse: Person-level cohort data (DCH–NG) and register linkages cannot be reused outside approved protocols or shared publicly. Any secondary analyses require new approvals from DCH–NG governance, ethics (if applicable), and updated DPAs. Trial person-level data and biobank samples may be used for related projects only with protocol amendments, ethics approval, and participant consent coverage. De-identified aggregate or derived datasets (where allowed) and analysis code may be shared openly after publication under approved licenses.
- Departing members: When a team member leaves the project or UCPH, their access to secure systems and sample facilities is revoked immediately. A handover plan transfers responsibilities, code, and documentation to the PI or designated analyst; all project data remain on institutional storage. Any personal copies must be deleted or returned, with confirmation logged. Authorship and IP follow collaboration agreements and journal policies.

3. Ethical and legal approvals

3.1 Describe any ethical considerations and approvals necessary for the collection, processing or use of material and data in your project.

- Cohort (Denmark): Use of DCH–NG baseline FFQ/lifestyle data with consented register linkage follows GDPR and Danish Data Protection law under DCH–NG governance. Analyses use pseudonymized files in secure environments. No new participant contact is required for this substudy.
- Randomized cross-over trial (Sweden): The Swedish team will obtain approval from the Swedish Ethical Review Authority and register the study (for example, ClinicalTrials.gov). Written informed consent will cover clinic procedures and coded storage of blood, saliva, and stool. Safety procedures include eligibility screening, adverse-event monitoring, and the right to withdraw. Samples are handled under Swedish biobank/SOP requirements (–80°C, key-file separation).
- Cross-border compliance: Any transfer of coded data or omics outputs between Sweden and Denmark will be governed by GDPR-compliant data-processing/sharing agreements specifying roles, security, and retention.
- Not applicable: No drugs/medical devices, gene technology, radiopharmaceuticals, animal experiments, or dual-use technologies are involved.

3.2 Describe any legal agreements or approvals necessary for the collection and use of material and data in your project.

- Personal data processing (UCPH): The cohort substudy and any Danish handling of trial data will be reported to the Faculty Secretariat using the UCPH “Form for reporting personal data processing in connection with research projects.” Processing follows GDPR and the Danish Data Protection Act with role-based access and DPAs where required.
- DCH–NG governance: Use of baseline FFQ/lifestyle data and consented register linkages is granted under DCH–NG data-use agreements; redistribution of person-level data is not permitted. Separate DPAs with register authorities govern linkage, retention, and disclosure control.
- Human participants, Sweden (trial): The Swedish Ethical Review Authority will approve the randomized cross-over trial, and participants will provide informed consent that covers processing of personal data and storage/use of biospecimens. Trial registration (e.g., ClinicalTrials.gov) will be completed.
- Biobanks and material transfers: Human biological material (blood, saliva, stool) will be stored under Swedish biobank governance. Any cross-border transfer of coded samples or raw omics data between Sweden and Denmark will use GDPR-compliant data-sharing/processing agreements specifying roles, security, and retention.
- Not applicable in this project: No import/export of live animals or cultural heritage objects; no use of genetic resources triggering ABS; no drugs/medical devices under the Danish Medicines Agency.

4. Collection, processing and documentation

4.1 Indicate what methods will be employed in the project to ensure the consistency and quality of the material and data.

4.1 Methods to ensure consistency and quality
Cohort (DCH–NG) data reuse and derivations

- Verification of source data: use only curated baseline FFQ/lifestyle datasets released by DCH–NG; run import checks (row/column counts, ranges, duplicates) and cross-check selected variables against codebooks.
- NOVA protocol: versioned, predefined decision rules; dual-coding of a sample of FFQ lines; adjudication log; sensitivity scenarios (lower/middle/upper bounds) to assess classification uncertainty; brief market audit of top SKUs to validate middle-bound assumptions.
- Organic indicators: documented mapping from FFQ questions to group-specific and overall scores; alternative codings pre-specified for robustness (e.g., binary vs ordinal).
- Analysis plan: preregistered SAP for the descriptive and association analyses; inclusion/exclusion criteria defined a priori; energy-intake plausibility checks; outlier rules documented.
- Bias control: consistent covariate definitions; missing-data strategy defined (complete case with imputation sensitivity); small-cell suppression in outputs.

Randomized cross-over trial

- Protocol and registration: ethics-approved protocol with SPIRIT checklist; trial registration (e.g., ClinicalTrials.gov).
- Randomization and allocation: computer-generated sequence (1:1 AB/BA), concealed allocation; prespecified handling of deviations; period/sequence terms in the model; carry-over tested.
- SOPs: standardized menus, weighing, and meal timing; pre-analytical SOPs for blood/saliva/stool (tube types, processing times, aliquoting, storage at –80°C); adherence checks and AE monitoring.
- Laboratory QA: instrument calibration and maintenance logs; internal standards and pooled QCs for LC-MS; sequencing controls (blanks, mock communities); randomized sample/batch order; replicate injections/runs as budget allows.
- Data capture: eCRF with range checks and audit trails; double-entry verification for critical fields.

Organization, documentation, and reproducibility

- Version control: Git repositories for code and protocols; semantic versioning of datasets (raw, cleaned, derived).
- Metadata: README files, data dictionaries, and flow diagrams (from raw to analysis tables); units and provenance recorded.
- Code review: internal peer review of analysis scripts; automated pipelines (e.g., Snakemake/Quarto) for end-to-end reproducibility; set seeds for random processes.
- QA audits: periodic QC reports (import checks, missingness, distributions, batch effects) archived as PDFs; governance log of decisions and deviations.

4.2 Describe how the material and data will be organized and structured.

Question not answered.

4.3 Describe how the collection, processing and analysis of the material and data will be documented.

Question not answered.

4.4 Describe what the approach will be for naming and versioning of data files and material.

Question not answered.

4.5 Indicate what metadata will be associated with the material and data.

Question not answered.

5. Storage and information security

5.1 Describe where and how the material and data will be stored and backed up during the project.

Question not answered.

5.2 Describe how the material and data will be shared with collaborators during the project (if any).

Question not answered.

5.3 For projects in which personal data are processed (including biobanks), please indicate whether a GDPR risk assessment and Data Protection Impact Assessment (DPIA) have been carried out.

Question not answered.

5.4 For all research data types, describe what security measures will be established to prevent breaches of confidentiality. How will unauthorized access be prevented?

Question not answered.

5.5 For all research data types, describe what security measures will be established to prevent loss of integrity. How will data and material be safeguarded against loss or modifications?

Question not answered.

5.6 For all research data types, describe what security measures will be established to prevent reduced availability of material and data. How will the continued accessibility of data and material to the relevant project members be assured?

Question not answered.

6. Data sharing

6.1 Describe which material and data will be made openly available for reuse.

- Protocols and documentation: NOVA coding protocol, decision rules, food-item dictionary, and analysis plans (PDF/Markdown).
- Code and workflows: scripts for data cleaning, NOVA assignment, exposure construction, and statistical analyses (Git repository; MIT/BSD license).
- Derived, de-identified aggregates: summary tables and figures (e.g., distributions of NOVA/organic intake overall and by sociodemographics), plus metadata/data dictionaries.
- QC summaries: non-sensitive quality reports (import checks, missingness, model diagnostics).

6.2 Are there any material or data that cannot be shared openly? Explain why.

Person-level cohort or trial data, raw omics files, and any files risking re-identification will not be shared openly; access to such data remains governed by DCH–NG and ethics/DPAs.

6.3 Describe any agreements required for sharing material and data.

Data Sharing agreement between Sweden and Denmark so that swedish trial data can be analysed in Denmark and combined with registry data.

6.4 *Findable:* What metadata will be created to allow the discovery of the material and data? How will others be able to discover the metadata?

- Metadata contents: Title, creators (with ORCID), contributors/affiliations, funder and grant ID, abstract, keywords (NOVA, ultra-processed, organic, DCH–NG), version, license, temporal/geographic coverage, methods summary, file formats, data dictionary link, and related DOIs (papers, protocol, code release).
- Documentation: Each record includes a README (scope, file inventory, provenance) and a data dictionary (variable names, codes, units).
- Persistent IDs and versions: DOIs for datasets and software releases (semantic versioning); link to exact Git tag/commit.
- Repositories and indexing: Deposit open materials (protocols, code, de-identified aggregates) in an institutional repository or Zenodo (DataCite-compliant). Records are indexed and discoverable via the repository, Google Scholar/Data, OpenAIRE, and linked from journal data-availability statements.
- Restricted data: For non-shareable person-level datasets, register a public metadata record stating scope, access conditions, and the DCH–NG contact/approval route.

Findable: **Will the material and data receive persistent identifiers?**

- Open materials (protocols, code, de-identified aggregates) will be deposited in a DataCite-compliant repository (e.g., Zenodo or institutional), which will mint DOIs per release.
- Authors will be linked via ORCID iDs; the trial will have a registry ID (e.g., ClinicalTrials.gov).
- All publications and data-availability statements will cite these persistent identifiers.

6.6 ***Accessible:*** **Where will the material and data be deposited (e.g. in which repository)? How will others be able to access and retrieve the material and data?**

Question not answered.

6.7 ***Accessible:*** **Will there be any conditions or restrictions for access to the material and data?**

Person-level cohort and trial data, raw omics files, and coded samples will not be openly available; access requires approvals from DCH–NG governance/ethics and GDPR-compliant data agreements. Public access will be provided to protocols, code, and de-identified aggregate summaries via an open repository under clear licenses.

6.8 ***Interoperable:*** **Will digital data be shared in file formats that others can easily open and reuse? Will information be provided on how files can be read and processed?**

Question not answered.

6.9 ***Interoperable:*** **Will standards for metadata (including vocabularies and ontologies) be applied?**

Yes. Public records will use DataCite/Dublin Core metadata with controlled vocabularies (e.g., MeSH terms for topics and FoodEx2/NOVA tags for diet exposures), and we will reference standard identifiers where relevant (ORCID for authors, trial registry IDs, and ontology terms from OBO where applicable).

6.10 ***Reusable:*** **What documentation is required for others to understand the material and data? How is this documentation provided?**

We will provide a complete package per dataset/code release: a README (scope, provenance, file inventory, versions), a data dictionary/codebook (variable names, labels, units, coding and missing-value rules), the NOVA protocol and decision log, methods/SAP excerpts describing exposure construction and models, and QC reports (import checks, distributions, diagnostics). Documentation is supplied as PDF/Markdown and linked to the exact code release (Git tag/DOI) in an open repository; licenses and access conditions are stated on the landing page.

6.11 ***Reusable:*** **Are there any conditions for the reuse of the material and data by others? How are these conditions communicated? Will you apply standard data usage licenses?**

- Open materials (protocols, documentation, code, de-identified aggregates) will be released under standard licenses. Reuse requires citation of the DOI, license terms, and acknowledgment of DCH–NG.
- Person-level cohort/trial data, raw omics, and coded samples are not available for open reuse. Access, if permitted, requires application to Sweden or DCH–NG governance/ethics and GDPR-compliant agreements.
- Conditions and licenses will be stated on the repository landing pages (DOIs) and in included LICENSE/README files. Any embargo periods will be indicated.

7. Long term preservation

7.1 Describe which material and data will be preserved after project end. Note that:

- **Unless regulated otherwise, research data underlying published results must be retained for at least five years.**
- **Projects may be required to register and possibly deposit research data with the Danish National Archives at project end.**

What will be kept

- Immutable copies of all datasets underlying published results (analysis-ready, pseudonymized), the exact code release (Git tag/DOI), and core documentation (NOVA protocol/decision log, data dictionaries, SAP excerpts, QC reports, READMEs). Retention at UCPH ≥5 years after each publication (longer if required).

Personal data and biospecimens

- Person-level cohort/trial data retained only in secure institutional storage under GDPR and cohort/ethics agreements; keep the minimal pseudonymized snapshots needed to verify publications or fully anonymize where feasible. Deposit with the Danish National Archives only if required; otherwise destroy/return per DCH–NG and approvals.
- Biospecimens stored or destroyed per Swedish biobank approvals; no reuse without new approvals.

Describe how the material and data will be preserved after project end:
Where will the material and data be stored? In which formats? For how long? What documentation and metadata will be associated?
As a minimum, describe how (a copy of) the digital data and corresponding documentation will be made available to research managers and/or supervisors at UCPH.

Question not answered.

7.3 Indicate who will have access to the material and data after project end.

If not otherwise discussed, then only the original working groups (Chalmers University in Sweden, and DCI in Denmark).

8. Resources and responsibilities

8.1 Describe what costs are associated with the management of material and data during the project. How will these costs be covered?

Question not answered.

8.2 Describe what costs are associated with the preservation of material and data after the project (according to the preservation plan and retention period outlined in question 7.2)? How will these costs be covered?

Question not answered.

8.3 Indicate who will carry out the different tasks for managing the material and data during the project, and for long term preservation.

Responsible person for registry data: Research group Diet, cancer, and health at the Danish Cancer Institute (group leader: Anja Olsen).
Responsible person for trial: Rickard Landberg at Chalmers University in Gothenburg, Sweden

8.4 Indicate who will be the main person(s) responsible during and after the project for:

- **Allocating costs and resources.**
- **Obtaining approvals and ethical assessments.**
- **Meeting legal and contractual obligations.**
- **Controlling access to material and data.**
- **Maintaining the integrity and availability of published and preserved material and data.**

Responsible person for registry data: Research group Diet, cancer, and health at the Danish Cancer Institute (group leader: Anja Olsen).
Responsible person for trial: Rickard Landberg at Chalmers University in Gothenburg, Sweden