DietWise

SYSTEMIC CHANGES | EMPOWERED CITIZENS

Deliverable D 1.1.

Data Management Plan 1

KU LEUVEN

https://www.dietwise.eu

Funded by the European Union

This project is funded by the European Union's *Horizon 2020 program* under grant agreement No. 101181692

Grant Agreement: 101181692| 2024 – 2027 | Duration: 36 months

Topic: HORIZON-CL6-2024-FARM2FORK-01-5. Type of Action: Research and Innovation Action (RIA)

Project information

Program:	Horizon Europe
Topic:	HORIZON-CL6-2024-FARM2FORK-01-5
Type of action:	HORIZON-RIA HORIZON Research and Innovation Actions
Grant Agreement #:	101181692
Project title:	DietWise: Systemic Solutions to Enhance Healthy and Sustainable Food Provision and Cooking at Home
Acronym:	DietWise
Project Start Date:	2024-11-01
Project End Date:	2027-10-31

Document information

Document name:	Data Management Plan 1
Related Work Package:	WP 1
Work Package Leader:	KU Leuven
Related Task:	Task 1.2 “Data Management”
Related Deliverables:	D 1.1
Author(s):	Ebo Botchway (KUL)
Reviewer(s):	Agnė Zakarevičiūtė (ADC); Siegfried Dewitte (KUL)
Submission date:	2025-02-28
Dissemination level:	Public

Document history

Version	Date	Changes	Responsible partner
v0.1	2025-01-31	First version of the document	KUL
v0.2	2025-02-24	Minor changes have been made regarding the content and the formatting	ADC
v0.3	2025-02-27	Final version of the document	KUL

Table of Contents

List of Tables

Abbreviations

	Deliverable
CATI	Computer-Assisted Telephone Interviewing
DCMI	Dublin Core Metadata Initiative
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findability, Accessibility, Interoperability, and Reusability
GDPR	General Data Protection Regulation
CATI	Computer-Assisted Telephone Interviewing
ICT	Information and Communication Technology
OSF	Open Science Framework
PIDs	Persistent Identifiers
SES	Socioeconomic Status

Executive Summary

The present document constitutes the initial version of the Data Management Plan (DMP) of the DietWise project, funded by the European Union's Horizon Europe Research and Innovation programme. It is designed to ensure proper, high-quality management of the data generated during the project. It describes the data generated during DietWise project activities, the potential for data exploitation beyond the end of the project, and the overarching data management strategy.

The DMP follows the template for "EU Grants: Data management plan (HE): V1.1 – 01.04.2022" and focuses on of implementing the FAIR principles for making research data findable, accessible, interoperable, and reusable. It also describes how open science practices are implemented as a cornerstone of DietWise data management to maximize the reliability, utility, and reuse potential of the data and other research outputs to be generated. The DMP specifies how DietWise data will be handled in a secure and ethically responsible way in line with the European Code of Conduct for Research Integrity and the General Data Protection Regulation.

The DMP is aligned with the DietWise Project Management Plan (Deliverable 1.1). Its implementation will be guided by the project's Data Manager from KU Leuven. This version of the DietWise Data Management Plan sets a basic framework of necessary actions, as perceived at this stage of the project. The DMP will be updated as additional details about the data to be generated during DietWise activities become known, and updated versions of the DMP will be published in months 18, 36, and 48.

1. About Dietwise

1.1 Project abstract

DietWise advances the state-of-the-art by proposing systemic changes, a focus on inclusion, and open social innovations with the aim of developing solutions that streamline existing tools and applications to foster healthy and sustainable food provision and to make cooking, eating, and treating of food at home the most attractive choice for all stakeholders. Using disruptive new approaches and voluntary market self-regulation, our activities will help to dampen nutritional noise gradually and organically merge cultural and commercial practices with a healthier and sustainable food consumption pattern. Next, we will empower citizens with novel, citizen science-based solutions that will shift the role citizens play – away from passive actors influenced by the food environment to citizens as active participants influencing their decisions and helping to create better digital food environments. The project will generate multiple novel insights to empower citizens to make healthy and sustainable choices and original empirical evidence through 'big data' analysis, large-scale surveys, qualitative research, and micro-level experiments. From a methodological point of view, we will harness novel state of the art methods, such as using artificial intelligence (AI) algorithms with deep learning techniques to analyze big online data. Finally, we will integrate scientific and empirical findings into social innovations powered by AI and develop user-oriented tools to support citizens. As a result, our methodological and empirical advancements will provide a deeper understanding of how various external system-level environmental factors shape attitudes and beliefs towards healthy and sustainable food provision and cooking, how to motivate consumers to follow nutrition guidelines, how to include the ones who are in greatest need, and how to help citizens shape digital food environments.

1.2 Consortium

DietWise is implemented by a well-balanced consortium comprised of high-level researchers in marketing, consumer behavior, psychology, nutrition, and information sciences, as well as industry experts. Our 10 partners represent three different countries (Belgium, Greece, and Lithuania). The consortium is a high-quality partnership based on diversified expertise, knowledge, and networks substantially covering all skills, resources, and access to stakeholders required for successful implementation. **All partners of the DietWise consortium adhere to sound data management principles** in order to ensure that the meaningful data collected, processed and/or generated throughout the duration of the project are well-managed, archived and preserved, in line with the Guidelines on Data Management in Horizon Europe and the General Data Protection Regulation (GDPR).

Table 1. DietWise consortium

No	Legal name	Short name	Country
1	KATHOLIEKE UNIVERSITEIT LEUVEN	KUL	BE
2	SAFE FOOD ADVOCACY EUROPE	SAFE	BE
3	FLEMISH INSTITUTE FOR HEALTHY LIVING (VLAAMS INSTITUUT GEZOND LEVEN VZW)	VIGL	BE
4	FOODBAG (SMARTMAT NV)	FOOD	BE
5	ADCOGITO, INSTITUTE FOR ADVANCED BEHAVIORAL RESEARCH	AdC	LT
6	VILNIUS UNIVERSITY FACULTY OF MEDICINE PUBLIC HEALTH DEPARTMENT	VU	LT
7	VILNIUS CITY MUNICIPAL PUBLIC HEALTH BUREAU	PHB	LT
8	INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	ICCS	GR
9	INTERNATIONAL HELLENIC UNIVERSITY	IHU	GR
10	PROLEPSIS	PROL	GR

2. Data Summary

Existing data re-use Tasks 2.1, 2.4, 3.1, 3.2 and 3.3 will utilise existing datasets to understand the current state of behavioral aspects associated with the uptake of nutrition guidelines, improvement of healthy and sustainable food provision, and eating and treating food at home. Other tasks will be guided by literature reviews of previous research but will address research questions that require the collection of new data. The initial scope of DietWise is set to collect, generate and re-use non-sensitive data that are not included into any special categories of personal data as those are described within the GDPR.

Types and formats of collected/generated data

DietWise combines qualitative and quantitative research methods.

Qualitative studies, such as interviews and focus groups, will include audio recordings and transcriptions of those sessions. Transcripts will be prepared in the language spoken during the interview or focus group, which could be English or the local language. The partner responsible for the qualitative research will also summarise the results in English. This textual data will be made available in openly accessible file formats like plain text, Unicode, and PDF. In addition, the text data

might be published in specific software formats, such as NVIVO, when such formats are likely to enhance data reuse.

Quantitative research methods—surveys, experiments, life cycle analyses, and data scraping—produce data in a tabular format and are typically found in spreadsheets. These spreadsheets detail the variations of one or more variables (e.g., responses to observed events, experimental stimuli, or survey questions) across different observation units, enabling quantitative analysis. The resulting spreadsheet data will be published in openly accessible file formats like CSV and PDF. Additionally, it may also be made available in proprietary formats specific to certain software (such as SPSS or Stata), supporting easier data reuse. Table 2 details the types and formats of the data to be collected and analysed within the DietWise tasks.

The purpose of the data generation or re-use and its relation to the objectives of the project

All data generation and re-use activities allow addressing research questions and/or evaluating innovation activities related to the DietWise project objectives:

O1. Understand food environment and citizen needs.
O2. Boost motivation to use nutrition guidelines for cooking with novel behavioral interventions.
O3. Develop digital social innovations and AI-based apps to empower citizens
O4. Empower vulnerable people to adopt healthy and sustainable food.
O5. Engage citizens to boost innovation through co-creation and citizen science.
O6. Pilot test the effectiveness of recipe apps and behavioral interventions.
O7. Build capabilities to follow nutrition guidelines.
O8. Develop a sample plan on how to enhance the uptake of beneficial tools and applications.

Table 2 details the specific objectives related to task-specific data generation and re-use activities.

What is the expected size of the data that you intend to generate or re-use?

Table 2 shows how the size of the dataset to be generated will differ across DietWise tasks.

What is the origin/provenance of the data, either generated or re-used?

Data will originate from existing databases, qualitative and quantitative research activities, and training evaluation focused on stakeholder capacity building (see Table 2).

To whom might your data be useful ('data utility'), outside your project?

Data-based reports will be useful to various stakeholders within and outside academia (e.g., influencers, retailers, consumers, policymakers). The datasets we generate will primarily interest other researchers. In addition, these datasets will be optimized for secondary research use before being published.

Table 2. Data to be generated or used through DietWise activities

WP2 Data collection/generation activities

Activity 1	**Institutional-level insights about key interventions to promote beneficial tools and applications uptake**
Purpose	WP2 focuses on providing institutional-level insights about key interventions to promote beneficial tools and applications uptake and compliance with the nutrition guidelines. In this task, the team will start with the inventory of key approaches used in the pilot countries, focusing on (a) existing nutrition guidelines promotion systems, processes, and schemes; (b) identification of existing tools and applications (c) factors that can facilitate or hinder the implementation of interventions planned by the DietWise project.
Task / Task Leader	2.1 / VU
Data	Information deriving from existing literature (published articles and reports), anonymized interview data
Data collection/generation process	Literature review, desk research, in-depth interviews
Data origin	Web sources and related publications, expert interviews
Storage format	Docx, pdf
Expected data size	Less than 1 GB
Other useful information	N/A
Activity 2	**A large survey on the needs, barriers, and drivers of citizens for uptaking beneficial tools and applications**
Purpose	The large survey in multiple EU countries will provide a deeper understanding of needs, barriers, and drivers influencing citizens to enhance healthy and sustainable food provision, eating and treating of food at home. In this task, the team will identify important factors - such as culinary culture, nationality, religion, culture, regionality, seasonality, food waste, and delivery, that may influence consumer decision-making, and cluster people based on this behavior. Apart from traditional demographic distinctions (e.g., age, gender), they will apply a psychological distinction between three large groups of populations, and study resistant, ambivalent, and supporting populations.
Task / Task Leader	2.2 / VU
Data	Survey data

Data collection/generation process	In this task, researchers will analyse the survey data to identify important factors that may influence consumer decision-making and cluster people based on this behavior. The data will be used to analyze the needs, barriers, and drivers of each target group. Depending on country specifics, the researchers will use the CATI (Computer-assisted telephone interviewing) technique, and/or online data panels, and/or crowdsourcing platforms (e.g., Sawtooth, Prolific) for data collection. To analyze variable relations, patterns, and groupings the team will use advanced inferential analytics (e.g., modeling, factor analysis, segmentation algorithms).
Data origin	Surveys
Storage format	Word (.docx), Excel (.xls, xlsx), csv
Expected data size	Less than 1 GB
Other useful information	N/A
Activity 3	**Interviews on how to boost the motivation of influencers to participate in voluntary initiative**
Purpose	This task will identify the underlying motivation of influencers, and define relevant themes, beliefs, and needs with regard to adhering to nutrition guidelines and participating in the initiative. The team will start by identifying what features of additional tools and apps may potentially boost and strengthen the business case of various influencers and how nutrition guidelines can provide added value to their followers.
Task / Task Leader	2.3 / SAFE
Data	Anonymized interview data
Data collection/generation process	Qualitative type of research- in-depth interviews with key influencers. The interview form will be prepared in English with a semi-structured, open-ended approach. If needed, these forms will be translated into local languages so the local coordinators can conduct the interview and provide a summary report in English
Data origin	Key influencers of the three pilot countries
Storage format	Word (.docx), Excel (.xls, xlsx)
Expected data size	Less than 1 GB
Other useful information	N/A
Activity 4	**Analysis of the needs, barriers, and drivers of vulnerable citizens**

Purpose	T2.4 aims to identify vulnerable groups and determine their current state of engagement with apps/tools/nutrition guidelines. We will start with desk research, a literature review, and at least five expert interviews per pilot country that will aim to define who are the most vulnerable citizens. Building on Task 2.2 results, the team will conduct a new survey targeting vulnerable citizens. The team will also analyze what factors are the most important for vulnerable citizens in uptaking of beneficial tools and applications (e.g., economic, financial, psychological, social, culinary culture, religious).
Task / Task Leader	2.4 / PROL
Data	Desk research, a literature review, expert interviews, and survey
Data collection/generation process	Web sources and related publications, semi-structured expert interviews. Depending on country specifics, the researchers will use available databases, and/or the CATI technique, and/or online data panels, and/or crowdsourcing platforms (e.g., Sawtooth, Prolific) for data collection.
Data origin	Literature review, desk research, anonymized in-depth interviews, survey
Storage format	Word (.docx), Excel (.xls, xlsx), csv
Expected data size	Less than 2 GB
Other useful information	N/A

WP3 Data collection/generation activities

Activity 1	**Behavioural data collection**
Purpose	WP3 aims to develop a package of the most efficient behavioral interventions by linking them to the role of nationality, religion, culture, regionality, and seasonality, reducing waste and delivery choices, as well as by matching the behavioral interventions with the needs of vulnerable citizens. Using an experimental approach, we will develop and validate the most efficient behavioural interventions.
Task / Task Leader	3.1/ KUL, 3.2. and 3.3 / ADC
Data	Controlled lab/online lab settings, literature review
Data collection/generation process	First, the team will conduct a brief literature review to identify potential behavioral interventions. Based on these insights and data from previous tasks (T2.1, T2.2, T.3.1and T3.2), the team will create experimental interventions. All interventions will first be tested in

	lab/online settings, and their effectiveness will be assessed based on the replicability of effects, effect sizes, and statistical significance. The data will be analyzed using analysis of variance and regression analysis.
Data origin	Anonymized data collected from professional research agency pools, crowdsourcing platforms, student samples in the lab or online settings
Storage format	Word (.docx), Excel (.xls, xlsx), csv
Expected data size	Less than 2 GB
Other useful information	N/A

WP4 Data collection/generation activities

Activity 1	**Co-creation of a framework addressing the needs of stakeholders**
Purpose	WP4 aims to co-create a framework for the DietWise project in line with the needs of relevant stakeholders, to create a corresponding capacity-building curriculum, as well as ICT solutions supporting this initiative
Task / Task Leader	4.1/ SAFE, 4.2 / ICCS
Data	Focus group discussions
Data collection/generation process	SAFE will organize co-creation workshops with influencers, citizens, vulnerable people, and other relevant stakeholders and gather responses from focus group questions about needs, barriers, and opportunities in T4.1 with the aim of increasing compliance with healthy and sustainable nutrition guidelines. Then, in T4.2, ICCS will organize focus group discussions with relevant stakeholders in ICT solutions and gather responses on needs, barriers, and opportunities with the aim of developing ICT solutions.
Data origin	Audio recordings and transcribed text documents from focus group discussions with influencers, citizens, vulnerable people, and stakeholders in ICT solutions.
Storage format	Word (.docx), Audio (.wma)
Expected data size	Less than 2 GB
Other useful information	N/A

WP7 Data collection/generation activities

Activity 1	**Pilot operationalisation in Belgium, Greece, and Lithuania**
Purpose	Task 7.1 aims to pilot test the effectiveness of different tools and applications in combination with the most efficient behavioural interventions and factors such as culture, religion, seasonality, and regionality on general population, vulnerable children, and adolescents.
Task / Task Leader	7.1 / VIGL
Data	Field experiments
Data collection/generation process	Using input from T5.3, we will gather observation records of app usage, choices, response times, accuracy in response to knowledge questions, stated and revealed preferences about healthy and sustainable food provision, eating and treatment of food, survey responses related to psychological characteristics of participants in Belgium.
Data origin	Anonymized data collected from citizens, vulnerable children, and adolescents
Storage format	Excel (.xls, xlsx), csv
Expected data size	Less than 1 GB
Other useful information	N/A
Activity 2	**Pilot operationalisation in Greece, and Lithuania**
Purpose	Task 7.2 aims to pilot test the effectiveness of different tools and applications in combination with the most efficient behavioural interventions and factors such as culture, religion, seasonality, and regionality on students from schools in SES neighborhoods
Task / Task Leader	7.2 / IHU
Data	Field experiments
Data collection/generation process	In Task 7.2, IHU will implement a pilot in Greece and gather observation records of app usage, choices, response times, accuracy in response to knowledge questions, stated and revealed preferences about healthy and sustainable food provision, eating and treatment of food, and survey responses related to psychological characteristics of participants in Greece, with the main focus on schools from SES neighborhoods.
Data origin	Anonymized data collected from students from schools in SES neighborhoods
Storage format	Excel (.xls, xlsx), csv

Expected data size	Less than 1 GB
Other useful information	N/A
Activity 3	**Pilot operationalisation in Lithuania**
Purpose	Task 7.3 aims to pilot-test the effectiveness of different tools and applications in combination with the most efficient behavioral interventions and factors such as culture, religion, seasonality, and regionality on vulnerable citizens.
Task / Task Leader	7.3 / PHB
Data	Field experiments
Data collection/generation process	In Task 7.3, PHB will implement a pilot in Lithuania and gather observation records of app usage, choices, response times, accuracy in response to knowledge questions, stated and revealed preferences about healthy and sustainable food provision, eating and treatment of food, and survey responses related to psychological characteristics of participants in Lithuania focussing on addressing the needs of vulnerable citizens.
Data origin	Anonymized data collected on vulnerable citizens
Storage format	Excel (.xls, xlsx) csv
Expected data size	Less than 1 GB
Other useful information	N/A
Activity 4	**Dynamic feedback, data collection, and analysis**
Purpose	WP7 aims to pilot-test the effectiveness of different tools and applications in combination with the most efficient behavioral interventions and factors such as culture, religion, seasonality, and regionality
Task / Task Leader	7.4 / ICCS
Data	User's perceptions
Data collection/generation process	ICCS will gather data on relevant feedback and feedbacks on online learning mechanisms. Also, observation records of app usage, choices, response times, accuracy in response to knowledge questions, stated and revealed preferences about healthy and sustainable food provision, eating and treatment of food, and user's preferences will be collected.
Data origin	Data collected through app usage
Storage format	Excel (.xls, xlsx), csv

Expected data size	Less than 1 GB
Other useful information	N/A

3. FAIR Data

DietWise partners will deposit all datasets suitable for publication in an open-science repository, such as Zenodo or OSF, that supports the FAIR principles by assigning persistent identifiers, creating metadata, and providing standard open-access reuse licences. Reports and publications based on these datasets will include links to the deposited datasets. While the datasets won't be published directly on the project website, the deposited datasets will provide links to the project's website. They will be published as soon as possible, and no later than the conclusion of the Work Package that produced the data or when any related report is published, whichever occurs first.

3.1 Making data findable, including provision for metadata

Will data be identified by a persistent identifier?

Each published dataset will receive a Digital Object Identifier (DOI) at Zenodo or the OSF. By employing Zenodo data repository, the data produced throughout the project is locatable by means of a standard identification mechanism. DietWise will be able to assign globally resolvable Persistent Identifiers (PIDs) on any data uploaded to Zenodo. In addition,Results not uploaded to Zenodo or another open access platform will be deposited on a searchable resource (e.g., the project website) and use tailored identification mechanisms—such as standard naming conventions—to ensure consistency and ease of discovery for project partners.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how?

Metadata will be created for each published dataset. Project data will be annotated with open and machine-readable metadata following the Dublin Core Metadata standard. The Dublin Core Metadata element set (certified with the ISO Standard 15836) is a standard that can be easily understood and implemented and, is one of the best-known metadata standards. It was originally developed as a core set of elements for describing the content of websites and enabling their search and retrieval. The Dublin Core metadata standard is a simple and effective set for creating rich metadata describing a wide range of resources. The fifteen-element "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The full set of vocabularies also includes sets of resource classes, vocabulary encoding schemes, and syntax encoding schemes. An online metadata generator will be used to produce the different metadata elements required (www.dublincoregenerator.com). For each published dataset, metadata will include at least the following information: title, description (including grant information), keywords, language, creation date, and modification date (if any). In addition, each published dataset will be accompanied by documentation (e.g., a codebook) that ensures an unambiguous understanding of variable names, labels, and values.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Metadata will include keywords that support the findability of published datasets.

Will metadata be offered in such a way that it can be harvested and indexed?

At both Zenodo and OSF (Open Science Framework), metadata associated with datasets that have been assigned a Digital Object Identifier (DOI) is structured and made accessible in a way that supports harvesting and indexing. This ensures that the metadata can be efficiently collected by external systems, such as search engines or aggregators, and made searchable across platforms.

3.2 Making data accessible

3.2.1 Repository

Will the data be deposited in a trusted repository?

Both Zenodo (dedicated repository for projects funded by the European Commission) and the OSF are generally trusted repositories.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Since publishing data and metadata are not linked to direct expenses at Zenodo and the OSF, and all DietWise partners can share datasets via those channels, we have not explored project-specific arrangements with the repositories.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Both Zenodo and the OSF allow datasets to be linked to DOIs, which resolve to the published dataset.

3.2.2 Data

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Generally, all datasets will be made publicly accessible. Before sharing, personal information within these datasets will be anonymized to ensure privacy. Project partners will implement appropriate steps to protect the identities of research participants and safeguard their personal information. In rare instances, like interview recordings, if data cannot be anonymized for public sharing, such data will be securely stored, with access granted to researchers only after obtaining the necessary ethical and legal approvals for the retention period. In these exceptional situations, DietWise partners should strive to provide the most comprehensive datasets possible while maintaining participant anonymity, such as through partially redacted interview transcripts.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No embargoes are anticipated. Datasets will be made openly available as soon as they are ready for publication.

Will the data be accessible through a free and standardized access protocol?

All data will be published on Zenodo or the OSF and, as such, they will be freely available for download via https.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project? How will the identity of the person accessing the data be ascertained? Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

As a rule, there are no restrictions on usage. Data that cannot be made openly available (e.g., interview recordings) will be securely stored, and researchers can be granted access to this data upon obtaining proper ethical and legal clearance for the duration of the retention period. The relevant ethical and legal authority will manage access to this data at the institution of the partner responsible for storing the data.

3.2.3 Metadata

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Metadata records will be openly available and CC0 licensed. Metadata for datasets with restricted access (if any) will include information on how to request access to the dataset.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Anonymized datasets and metadata will be permanently available on the open-science repositories, and they will be retained for the lifetime of the respective repository.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

All published datasets will be provided in openly accessible file formats. Additionally, data records may feature datasets in software-specific formats (e.g., SPSS, Stata) to enhance data reuse. If available, processing and analysis scripts or code will accompany the published datasets.

3.2.4 Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

Project partners will guarantee the interoperability of datasets by following common practices in their respective fields. Typically, this will involve 1) using openly accessible standard file formats, 2) providing data in both unprocessed and processed forms, 3) supplementing datasets with processing and analysis scripts, 4) employing standard table formats and ensuring consistent formatting of cell values (for spreadsheet data), and 5) augmenting datasets with documentation (e.g., codebooks) to enable unequivocal understanding of variable names, labels, and values.

The interoperability of data that will not be publicly shared will be ensured through the Dublin Core Metadata standard**.** This standard is a small "metadata element set" which accounts for issues that must be resolved in order to ensure that data meet traditional standards for quality and consistency, while remaining broadly interoperable with other data sources in the linked data environment. The standard includes fifteen elements that define key concepts, such as title, creator, and author, using natural-language descriptions that are automatically converted into open machine-readable formats like XML and HTML. Each element is optional and can be used multiple times, with built-in mechanisms for refinement, supporting the use of encoding and vocabulary schemes.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Due to the absence of commonly used data and metadata ontologies in the disciplines contributing to DietWise, mapping DietWise data and metadata to these ontologies is not feasible. Published datasets will include openly available documentation that guarantees an unambiguous understanding of variable names, labels, and values.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

Where applicable, metadata will include qualified references to the persistent identifiers of additional datasets that complement the dataset at hand. These references will specify the type of relation between the datasets.

3.2.5 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Alongside each published DietWise dataset, corresponding documentation will be made available in the same trusted repository. This documentation will include a file (in an openly accessible format) that explains all variable names, labels, and values included in the accompanying dataset (for instance, a codebook). Where applicable, processed datasets will also be provided alongside unprocessed datasets, processing scripts, and readme files that detail how to apply the processing scripts to transform unprocessed data into processed data. Furthermore, analysis scripts will be made openly available, and readme files will indicate how those scripts should be used to reproduce the results we obtained from the published data. When the results of DietWise data are published in scientific journals, this documentation may also be published as supplementary material alongside the journal article.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Each published dataset will be made available under a Creative Commons Attribution Required CC-BY 4.0 licence. For datasets that cannot be published in their entirety (e.g., because data cannot be anonymized for open sharing), DietWise partners will make reasonable efforts to publish (under CC-BY 4.0 licence) datasets that are as rich as possible while ensuring participants' anonymity.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

All data deposited in online repositories will be usable by third parties (with attribution required) after the end of the project. Third parties can also obtain access to unpublished datasets (provided proper ethical and legal clearance) for the duration of the retention period.

Will the provenance of the data be thoroughly documented using the appropriate standards?

Each published dataset will include documentation of data provenance and links to any reports detailing the data collection process.

Describe all relevant data quality assurance processes.

DietWise partners will adhere to best practices within their respective research disciplines regarding the collection, processing, analysis, sharing, and documentation of data. Where applicable, researchers will preregister data collection activities and utilise established (or validated) procedures to obtain valid, reliable, and useful results that benefit other researchers.

Each data collection activity will be coordinated within Work Package Teams. The DietWise Coordination Team will provide input on data collection activities and approve all deliverables.

Guidelines for the FAIR sharing of datasets will be distributed to all partners by the DietWise Data Manager, and the Data Manager will review all datasets upon repository deposition.

4. Other Research Outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

The research efforts of DietWise will generate data processing and analysis scripts that will be deposited along with the corresponding datasets in dependable repositories. Each script will have annotations to support reuse and will include readme files detailing their use. Additionally, DietWise will publish preregistered study protocols under standard reuse licenses (e.g., on OSF). Furthermore,

the stimulus materials, behavioral experimental procedures, and metrics created for or utilised in DietWise research will be made publicly accessible (in online repositories or as supplementary materials with journal articles) whenever legally permissible.

5. Allocation of Resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

Data, scripts, preregistrations, and research materials can be deposited and published in Zenodo and the OSF free of charge.

Any costs related to using (secure) infrastructure for collecting, processing, storing, and exchanging data during the project are covered by the host institution of the respective DietWise partner. DietWise includes a dedicated task to coordinate and guide the implementation of FAIR data management throughout the project.

Who will be responsible for data management in your project?

DietWise Work Package and Task leaders are responsible for ensuring that all data are collected and managed in accordance with the Data Management Plan and the Project Management Plan, as well as institutional and national guidelines and policies. This includes securing clearance from institutional review boards or ethics committees before initiating data collection.

The DietWise Data Manager at KU Leuven (Ebo Botchway, ebo.botchway@kuleuven.be) will guide the implementation of the Data Management Plan, support project partners in establishing FAIR data sharing procedures, and review datasets before publication. The Coordination Team will advise in case of doubt regarding project-related data and ethics management.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Deposited data, along with the accompanying metadata, documentation, and other research outputs, will remain accessible for the lifetime of the repository. Retention periods for data that cannot be publicly deposited (e.g., personal data) will adhere to relevant national and institutional regulations and guidelines. Project partners will ensure that the retention of such data aligns with approvals from the appropriate data protection authorities (institutional review boards, ethics committees). Partner institutions will cover any additional costs associated with long-term storage.

6. Data Security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

The collection and management of personal data will comply with the General Data Protection Regulation (GDPR). To ensure that data management conforms to the relevant institutional and national regulations and guidelines, approval from institutional review boards or ethics committees will be obtained.

The collection of personal will be restricted to what is strictly necessary for the intended purposes, preventing unnecessary data accumulation and maintaining data relevance. To collect and store personal data, secure platforms will be utilised that can only be accessed using multi-factor authentication. For storage, datasets will either be anonymised (i.e., through the deletion of personal data that could identify individuals) or pseudonymised and encrypted/password-protected. Data will be stored on secure servers that are regularly backed up.

Before sharing datasets between institutions, data will be anonymised, or the institutions will formulate data transfer agreements for the exchange of personal data (in accordance with relevant procedures at the respective institutions). Anonymised datasets and other forms of non-sensitive data will be exchanged among project partners using the DietWise SharePoint platform at KU Leuven. This solution provides access control, backup procedures, and file versioning.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Anonymised datasets, along with the accompanying metadata, documentation, and other research outputs, will be permanently stored in trusted repositories (e.g., Zenodo, OSF) for the lifetime of the repository.

7. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

As specified in Section 6 of the Project Management Plan (Deliverable 1.1), DietWise partners will conduct thorough cost-benefit analyses to ensure that the societal and scientific benefits of research activities surpass any potential risks to participants. Throughout all research activities, they will aim to maximize these benefits (e.g., by collecting robust data and facilitating data reuse) and minimize risks (e.g., limiting personal data collection and safeguarding participants' privacy). Institutional review boards or ethics committees will review all planned research activities and the associated ethical cost-benefit analyses, and data collection activities will only commence once ethics approval has been obtained.

Research participants will be informed that their data will be made publicly available in anonymized form in a trusted repository.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

All DietWise research activities that involve collecting data from identifiable research participants will obtain informed consent. Consent forms will inform participants that their data will be made permanently publicly available in anonymized form in a trusted repository. The precise form and content of the informed consent form will depend on the local ethics authority.

8. Other Issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

DietWise partners will consider the Guidelines on FAIR Data Management in Horizon 2020 and the European Code of Conduct for Research Integrity to maintain ethical research practices in all activities, along with any institutional guidelines concerning research integrity, open science, and data protection at their respective partner institutions.