EMPOWERING HEALTHY LIFESTYLE BEHAVIOUR THROUGH PERSONALISED INTERVENTION PORTFOLIOS TO PREVENT AND CONTROL OBESITY DURING VULNERABLE STAGES OF LIFE

Grant Agreement Number 101080645

Deliverable number	D7.2
Deliverable title	HealthyW8 Data Management Plan (DMP)
Deliverable type:	R
Work Package:	WP7: Project management and coordination
Lead beneficiary:	LIH
Contact person:	Torsten Bohn: torsten.bohn@lih.lu
Contributors:	LIH : Torsten Bohn, Michel Vaillant, Amir Sharify, Laurent Prévotat, Yacine Ouzzahra, Mariana Alpalhão Gonçalves TU/e : Pieter Van Gorp, Edyta Cios
Dissemination Level:	PU: Public
Due date for deliverable:	M6
Actual submission date	M8
Revision number	V1.0

ACKNOWLEDGEMENTS

The research leading to these results has been conducted as part of the HealthyW8 project, which received funding from the European Union's Horizon Europe Research and Innovation Programme under the grant agreement n°101080645.

EXECUTIVE SUMMARY

Background	The Data Management Plan (DMP) describes all the data management processes related to the HealthyW8 project. Careful versioning of the DMP will ensure traceability of the changes and contribute to the FAIR nature of data and metadata.
Objectives	The DMP identifies the types of data that will be collected and processed within the project and how this data will be handled during and after the project.
Methods	This DMP has been prepared using the "Horizon Europe Data Management Plan template" (version 1.0, 2021).
Results & implications	This document has been divided into four sections with additional annexes: • Data Summary • FAIR data • Allocation of resources • Data security • Ethics and Privacy **It is important to note that some questions concerning HealthyW8 data are still open for discussion at this stage of the project. This deliverable is therefore a living document and will be updated throughout the project lifecycle. At month 30 (October 2025), an updated version (V 2.0) of the DMP will be submitted (deliverable 7.3)**

Table of contents

Funded by
the European Union

1 Introduction

1.1 HealthyW8

Over 30% of EU citizens at vulnerable stages and situations in life are at increased risk to cross over from healthy weight to overweight and further to obesity. Though many interventions have been proposed to tackle obesity, they have rarely been effective. The aim of HealthyW8 is to advance the efficacy of current and future efforts and investments in obesity prevention initiatives across Europe. Most interventions focus only on diet and/or physical activity while overlooking emotional and user motivation aspects. The lack of adaptation to personal contexts such as socioeconomic facets, host-biological factors, environment, dietary preferences or fitness level often leads to initiatives on obesity prevention only having a marginal impact. HealthyW8 will address these shortcomings by iteratively developing, together with stakeholders, a digital-based healthy lifestyle recommender (HLRS) for evidence-based, tailored interventions and tools including a human digital twin (HDT) to bridge the gap between science, societal actors and stakeholders (e.g. healthcare professionals, food industries, policymakers) and EU citizens. The targeted populations are those undergoing transitions, i.e. schoolchildren (5-12 y, and their parents), young adults (18-25 y) and the elderly (≥65 y). In the mid-term, we estimate that with 200,000 HealthyW8 users, we will prevent 10,000 obesity cases/year. In the long run, the impact will be maximized through adopting the project's proposed methodology, platform and tools by as many EU institutions and entities as possible. HealthyW8 is a highly experienced, synergistic and complementary consortium that will build on a previously developed digital dietary app (LIFANA) and draw on transdisciplinary research in pan-EU multicentre pilots and long-term randomized control trials to achieve its overarching objective of increasing impact of current and future interventions on obesity prevention and policies in the EU.

1.2 Project Management and Coordination

One dedicated workpackage of HealthyW8 (WP7) sets forth a three-pronged strategy composed of the following: (1) strategic steering of the project and addressing all unexpected situations; (2) operational monitoring and risk assessment to ensure progress in conformity with the work plan regarding overall work, milestones, deliverables and use of resources, as well as to survey ethical and gender issues; and (3) administrative, financial and risk management to ensure day-to-day coordination and consortium facilitation. We will also guarantee IPR management according to the EC Guidelines, the Consortium Agreement. WP7 will be flexible to enable the

project to remain adaptable to external events, to ensure the project progresses in conformity with the work plan, and transparent to create trust between the partners.

Within this workpackage, one task is dedicated to the coordination of knowledge management, an consists in elaborating a data management plan (DMP) ensuring that the consortium will manage the data responsibly and in line with FAIR principles, as well as managing the IP of related to the project's results. During the project, potentially valuable IP will be captured, evaluated and protected by the appropriate means. Upon consultation with the executive committee (ExC), we will ensure that dissemination activities do not jeopardize market entry and will verify that sensitive information is adequately protected. In case of co-ownership issues between different partners having different policies, solutions will be proposed to the concerned partners.

2 Data Management Plan (DMP)

2.1 Data summary

The core objective of HealthyW8 is to develop, evaluate and optimise user-centred multi-level interventions to reduce the risk of transitioning from excess weight to obesity, in 3 different age-groups across Europe. To this end, one of the project's central pillars will be a series of multicentre pilots (≤ 3 months) and long-term (12 months) randomised controlled trials. Data collection, storage and processing are therefore crucial aspects of the project. Table 1 below provides an overview of modules (i.e. categories) of data identified for the HealthyW8 project, as well as the corresponding measurement tools and specific parameters.

Module	Target	Tools	Parameters (examplatory list)
Baseline and endline survey	Biological sample characterisation Socio-demographic data Quality of life Mental health	Questionnaires to be implemented into an EDC system (e.g. RedCap): - WHOQOL-BREF - KIDSCREEN-27 - SDQ - PANAS	Age, gender, medical history, quality of life, positive and negative affect, depression, alcohol, nicotine
	Anthropometry	Measurement by nurse	Height, weight, BMI, waist-hip ratio, thigh circumference, visceral fat (calculated)

(Bio-) marker assessment	General health aspects	plasma analysis, bio-impedance	Blood glucose, insulin, HbA1C, leptin, HOMA-IR, blood lipids, pulse, blood-pressure, %age body fat, blood cell counts
	Inflammation	Plasma analysis (ELISA...)	Cytokines, CRP
	Oxidative stress	Plasma analysis (ELISA...)	F2 isoprostanes, malon-dialdehyde, antioxidant enzymes, DNA/RNA breakdown, antioxidant capacity
	Nutritional	Plasma and urinary analyses	Vitamin E, carotenoids, creatinine, total phenolics, uric acid, vitamin D
	(epi-)genetics	Blood cell analyses	SNPs*, DNA methylation*,
	-omics	Saliva and plasma analyses	Saliva proteomics, microRNAs
	Gut microbiota	Fecal sample analyses	16SrRNA, metagenomics*
Dietary & general counselling	Improved dietary habits, adaptation to individual	Food frequency questionnaires (FFQ), consultation with nurse/dietician & adapting the meal recommender system	Macro-, micro-nutrients, non-nutrients, healthy life-style questionnaire, questionnaire on sleeping patterns and psychological state
PA	Altered physical activity (PA) pattern	Actigraph (research grade accelerometer)	Changes of PA (light PA, MVPA, distribution of intensities) and SB (total time, bout duration and distribution/accumulation);
HLRS – use of functionalities	Healthy and balanced personalized diet	Personalized meal recommender plan (based on age, gender, physical activity (PA), dietary restrictions, culinary/ cultural personal preferences & local dishes, budget…)	Local adapted meals proposed, assessment of cooking and food purchasing patterns
	Stimulation of PA & reduce sedentary behavior (SB), adapted to individual abilities & fitness	Motivational features	Nudging, gamification, recommendations by HLRS & nurses, suggestions to join sport clubs, etc.
	Measured PA and SB	Smartwatch or Fitbit, to be determined in co-creation & evaluated in pilots	Long-term determination of active time, step counting etc. (Fibit...),

	Feedback on emotions / behavioural / psychological aspects	Multi-modal signal analysis tool, app with choosable emoticons, e-questionnaires	Heart-rate variability, emotional state, Minimental health questionnaire
	Healthy, individualized sleep patterns	App recommendations based on individual lifestyle, sleep features (accelerometer)	Total sleep duration, time in bed, sleep efficiency
	Giving psychological support	App assessing emotions and giving feedback	Stigmatization mitigation, advice on when to seek professional health
	Assessment of physical condition & stress	Multi-modal physiological signal analysis	Heart rate & variability, blood pressure, PA, skin conductivity etc.
	Improved & sustained adherence/motivation to use the HLRS	Motivational features	Nudges/gamification, quizzes, virtual prices, defaults, information, presentation, priming
	Enhanced social interactions	app recommendations	Social activities, support from peers, family etc.
	Stimulation of consumption of more sustainable food	Link to food product or life-cycle (LCA) databases (Agribalyse or other	Sustainability indices (CO_2 equivalents, water consumption, etc.).
HLRS - telemetry	Telemetry data: statistics on usage of the respective Apps (and functionalities) of the HLRS	Applications	Usage (e.g. duration, time of day, frequency), settings implemented (e.g. silenced notifications), time spent on specific functionalities of the HLRS, etc.
HDT	Improvement of lifestyle aspects	Emotions-aware HDT predicts barriers to healthy lifestyle	Emotions, psychological state, socio-demographic aspects
	Upskilling of health literacy	Assessment and learning	Individual competencies
Policies, recommen–dations	Stimulating healthy living environment	Local or nationwide policies, recommendations, guidelines	Suggestions on diet, general healthy lifestyle (e.g. sleep patterns), taxation of unhealthy foods, healthcare system incentives, socio-economic aspects etc.
OSP	Optional, additional healthy lifestyle recommendations and educational aspects	Health-clips, recipes, lifestyle recommendations	Information on facilities that support healthy behaviour (parks, clubs), targeting diet, PA, healthy lifestyle

Table 1: *Multi-disciplinary portfolio intervention strategy and parameters assessed*

Data will be pseudonymised and any identification data (e.g. names, contact details) of all participants involved in the project will be kept separate from experimental data and be accessible using security passwords. Electronic personal datasets will comply with national legislations. These identifying data will not be exchanged between partners.

The only secondary data evaluation that will take place in the project relates to the systematic reviews and scoping reviews currently undertaken as part of WP1.

2.2 FAIR data

2.2.1 Making data findable, including provisions for metadata

To ensure findability of the data, all published datasets will be given a persistent identifier/digital object identifier (DOI), created as part of the data publishing process on the data repository.

To allow the discoverability of the data, the persistent identifiers of the datasets will be (hyper)linked to the journal (or other) publications.

In addition, all datasets are described by the metadata and documentation attached to them during publication.

Pseudonymized research data will be initially collected and stored in a federated manner by the partners using relevant open, lossless formats, which will later facilitate recombination of datasets from different origins.

2.2.2 Making data accessible

Only fully anonymized datasets will be published. We will always take the necessary steps to delete any identifying information from the datasets.

Some outcomes of the project will lead to patent applications. How we deal with ownership of data and intellectual property rights (IPR) has been laid down in a specific Consortium Agreement for this project, agreed upon by all project partners. IPR protection will be first evaluated for each of the outputs (including data).

All partners will be able to access project data for at least 5 years after its end. Meta-data will be provided in order to ensure interoperability for indexing and discoverability.

Provided that adequate consent was obtained from study participants and HealthyW8 partners, these data can be reused, beyond the project, by all partners, for pursuing additional research questions or generating novel hypotheses for further publications or project applications.

Data collection as part of the pilot trials will be done through both centralised and local data collection systems, while a unique centralised system will be used especially for the long-term trials. Other data, collected from stakeholders or from literature review, will also be stored in a shared, password-protected domain.

A first estimate of the required space for data storage was performed, indicating a range between 50 and 100 TB, for a maximal duration of 15 years (5 years project + 10 years storage of raw data required by the E.C.).

The Data Integration and Analysis department of LIH (DIA) will provide the needed IT infrastructure for data storage in Luxembourg as well as later on setting up the data glossary in order to facilitate the data reuse after its storage in a centralized data warehouse. The storage provided will increase gradually during the enrolment of the study in order to satisfy the required capacity in order to host all types of data that will be collected in the framework of this project.

The centralized research data will be stored in two geographically separated LIH data centres. The data are stored in redundant fashion on multiple servers. The servers are mirrored in both data centres.

The data will also get fully backed up regularly. The backup servers are situated in a physically secured compartments. Recovery tests are performed in regular intervals. The backup process is governed by Standard Operating Procedures.

2.2.3 Making data interoperable

Standard procedures will be used in all steps of the data collection to ensure standardisation, data quality and data integrity. Coding instructions (e.g. Data Entry Guideline, etc.) will be provided for the eCRF data as well as any other data collection tool provided to patients.

To ensure automatic findability and interoperability of datasets, commonly used controlled vocabularies, ontologies and thesauri will be used to describe them. Data coding standards are important as they are key to semantic interoperability, i.e. the possibility to use and exchange data for different purposes while retaining the true meaning of the information. These standards can be later mapped into other standard terminologies, facilitating the data exchange process.

Meta-data will be defined based on medical informatics standards per data categories. All data elements will be

annotated in a common data catalogue. This data catalogue contains annotation of all data elements and references to the available datasets.

2.2.4 Increase data re-use

The data transformation will be supervised by data managers. The collected data as well as any output-curated data are checked via automated and manual unit tests (e.g. source data verification and automated edit checks). For the data mapping to the common interoperable data standard between sites a harmonized implementation guide will be used. The common data model (core set of variables) will use ontologies and standards.

All the different raw datasets will be collected in a data lake architecture. The raw datasets will be exported into the data lake in their native formats (e.g. eCRF as CDISC ODM).

To increase transparency and accessibility to the evidence generated by this research, project partners commit to the Gold Open Access route to scientific publications and to open access to research data (anonymised as needed in accordance with ethical and governance requirements due to their sensitive nature) and other materials and analysis codes (e.g. software, models, algorithms, workflows). The publication in Open Research Europe will also be targeted. In a quest for the highest impact of our results, specific hybrid journals with a high impact in their fields may also be considered for publication (in this case, the funds for the APC fees will be obtained from other sources). In parallel, publications and other outputs generated by the project will be archived via the in an Open-Access repository. Each partner will also provide, via the repository, information about any research output, tool or instrument needed to validate the conclusions of the publication to enable reproducibility.

2.3 Allocation of resources

Costs related to open access are eligible as part of the Grant, and therefore foreseen in the HealthyW8 budget including legal fees involved in ensuring the data sharing and the data repository costs. Still, some considerations to think about are the resources for long term preservation of the individual data sources included the repository, at what costs can the repository operate during and after the project, and for how long.

Several HealthyW8 partners have dedicated support staff for achieving Open Access. Additionally, support is also available on topics such as research data management, legal counsel, and privacy.

Making the data FAIR (assigning DOI, describing the data, assigning metadata to the data, etcetera) is part of the ongoing research activities of the involved researchers.

2.4 Data security

The minimal data to be collected will be specified within the ethics protocol and the corresponding questionnaires for collection will be developed. Privacy by design, data minimization and pseudonymization will be key principles to be adhered to, throughout all the phases from system design, data collection and transfer as well as data storage and reporting.

Data will be collected by using as a GCP and GDPR compliant data management system. Procedures will be developed on how to transfer the individual data sources that are pseudonymized to the repository in a standardized and harmonized way.

All these data will be under the responsibility of the Principal Investigator (PI) of the individual dataset, and the PI will have the duty to curate and pseudonymize the data. All personal data is stored in secure and dedicated databases, which will be preserved in compliance with the European GDPR and national data protection legislation.

Pseudonymized data will be stored in a data repository, and data will be treated according to the minimization, accuracy and storage limitation principle, as stated in the General Data Protection Regulation (GDPR). Access to the repository and its underlying data will be regulated via a rights and roles management system guaranteeing data security. In the project's lifetime, only partners involved in the consortium will have access to the data repository, and options for operating the data repository post- project-life should still be discussed with all partners involved.

Privacy will also be protected when results or data are presented, i.e. as a general rule, restrict to aggregations or line listing deprived of personal identifier to prevent re-identification.

2.5 Ethics and Privacy

The Cohort Owner acting as the study participant's interlocutor is responsible for maintaining at all times proper documentation of Ethics approval. Each Cohort Provider warrants that:

- The Ethics approval and Informed consent:
  - ❖ has been obtained from all concerned study participants, and is properly documented.
  - ❖ covers all Data being processed under or in connection with the Research Project
  - ❖ complies with, and/or includes the information required under articles 12 to 14 GDPR[1],
- No local laws applicable to Cohort Provider are prohibiting the use of Data processed during the project,
- As applicable, no DPIA (Data Protection Impact Assessment) is required for the Processing, or a DPIA has been conducted and its conclusions (including Supervisory Authority's consultation as potentially required) are positive and allow the Processing to be carried out in the conditions of the Research project.

the collection and onwards processing of Data is grounded on a lawful basis, as required under the GDPR.

Each Cohort owner is responsible for handling Data Subjects' queries, claims or requests tending to the exercise of their statutory rights in the context of the Processing.

The consortium's objective is to collect high-quality standardized data, and to achieve this, necessary ethical approval will be obtained prior to the launch of the studies.

Personal data are only processed if the written informed consent of the study participant is given.

In addition to compliance with the applicable legal framework (GDPR and research), data protection, cybersecurity and transparency are the core values of the HealthyW8 consortium.

In particular, principles of data protection by design and by default will be applied. Non-exhaustively, the consortium should:

- Ensure the possibility to exercise the data subjects' rights and ensure alignment with all partners. In particular, participants will have the right to access, edit and suppress their data, upon request.
- Ensure data-level encryption whenever possible as well as encryption of databases and hardware.
- Provide strict user and access management with role-based and person-based access rules as well as a high-standard authentication and authorisation infrastructure.
- Limit access to subsets of the data (by patient; by data type).

[1] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

- Ensure logging of server operations at all times.
- Provide a data protection concept with distributed data holding, trusted third party for initial pseudonymisation and concepts for re-pseudonymisation where data are used in projects or shared with others.
- Provide a consent management system to build automatic purpose limitation compliance.
- Enable a data information system to store or link to all GDPR-relevant metadata.
- Provide granular user access to different databases

Transparent and targeted communication with participants and key stakeholders will be of the utmost importance to create a trust relationship.