Youth health from a holistic perspective

Grant Agreement 101155977

Deliverable 11.3
Data Management Plan - Initial

Version 3.2 | June 2025

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Deliverable Title	Data Management Plan - Initial
Brief Description	Plan providing guidance on how to deposit data in a clear and transparent manner.
Work Package No	11
Deliverable No	11.3
Lead Beneficiary	1 - UGA
Authors	Laetitia Minniti (UGA), Aïna Chalabaev (UGA), Sandrine Isoard-Gautheur (UGA), Antoine Marin (UR2), Alexa Comte (UGA), Ann DeSmet (ULB), Sandrine Isoard-Gautheur (UGA), Sebastien Chastin (GCU), Giuliana Longworth (GCU), Anca Dobrean (UBB), Vincent Brugger (UHEI), Nerea Moreno (ISGlobal), Carmen Peuters (ISGlobal), Payam Dadvand (ISGlobal), Daniel Phipps (JYU), Boris Cheval (ENS Rennes), Paulina Alicja Jedynak (IISPV), Jordi Julvez (IISPV), Fabienne Fox (Accare)
Deliverable Due Date	June 30, 2025
Actual Delivery Date	June 26, 2025
Nature of the Deliverable	DMP
Dissemination Level	PU - Public

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Document History

Date	Version	Contributors	Comment
09/04/2025	1.0	Laetitia Minniti (UGA)	First draft
23/04/2025	1.1	Aïna Chalabaev (UGA)	
24/04/2025	2.0	Laetitia Minniti (UGA)	
24/04/2025	2.1	Laetitia Minniti (UGA)	
28/05/2022	3.0	Ann DeSmet (ULB), Sandrine Isoard-Gautheur (UGA), Antoine Marin (Rennes2), Alexa Comte (UGA), Sebastien Chastin (GCU), Anca Dobrean (UBB), Vincent Brugger (UHEI), Nerea Moreno (ISGlobal), Carmen Peuters (ISGlobal), Payam Dadvand (ISGlobal), Daniel Phipps (JYU), Boris Cheval (ENS Rennes), Paulina Alicja Jedynak (IISPV), Jordi Julvez (IISPV), Fabienne Fox (Accare)	Additional information from the WPs
05/06/2025	3.1	Aïna Chalabaev	
17/06/2025	3.2	Giuliana Longworth (GCU), Alexa Comte (UGA), Sandrine Isoard-Gautheur (UGA), Antoine Marin (Rennes2), Anca Dobrean (UBB), Vincent Brugger (UHEI)	Additional information from the WPs

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Executive Summary

This Data Management Plan (DMP) outlines the strategy for handling data throughout the lifecycle of the YEAH! project, in alignment with Horizon Europe requirements and **FAIR** (Findable, Accessible, Interoperable, Reusable) data principles.

The DMP first presents an **overview of the project's objectives** and explains how data generated or reused contributes to achieving them. It details the types of data expected, formats, origins, volume, and intended audiences beyond the project. Data will be processed with appropriate tools and resources, and partners will ensure that **metadata** and documentation enable future validation and reuse.

Clear **guidelines** are provided for ensuring that data are findable and accessible, with the use of **persistent identifiers** (DOIs), standard **naming conventions**, and deposition in trusted **repositories** (e.g. Zenodo). Metadata will be published under a CC0 license and include all necessary identifiers and descriptive elements, as stipulated in the Grant Agreement (GA).

Provisions for data **interoperability, reusability, and long-term preservation** are also addressed. While some processes are yet to be fully defined, the DMP proposes initial strategies that will be refined as the project progresses.

Legal and ethical considerations, including compliance with the General Data Protection Regulation (GDPR), intellectual property management through the Consortium and Grant Agreements, and ethical oversight via a dedicated Ethics Management Plan and Ethics Advisory Board, are described.

Finally, the plan includes details on **roles and responsibilities**, associated **costs**, data security and backup procedures, and annexes documenting project datasets and templates.

This DMP is a **living document** that will evolve during the project, ensuring that the data strategy remains aligned with scientific goals, legal requirements, and open science principles.

Any new beneficiary, partner, or affiliated entity joining the project is required to **comply with the DMP**. It will be updated in a timely manner to reflect such changes and ensure that all data management obligations are clearly communicated and followed.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Overview

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Abbreviations

ABCD	Amsterdam Born Children and their Development
BiB	Born in Bradford
BREATHE	BRain dEvelopment and Air polluTion ultrafine particles in scHool children
CA	Consortium Agreement
CC BY	Creative Commons Attribution International Public License
DCC	Digital Curation Centre
DTA	Data Transfer Agreement
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
EC	European Commission
ExCom	Executive Committee
FAIR	Findability, Accessibility, Interoperability, Reusability
GA	Grant Agreement
GDPR	General Data Protection Regulation
HE	Horizon Europe
IPR	Intellectual Property Right
ISSN	International Standard Serial Number
NDA	Non-Disclosure Agreement
NGO	Non-Governmental Organization
OA	Open Access
PII	Publisher Item Identifier
PMP	Project Management Plan
PMT	Project Management Team
RCT	Randomized Controlled Trial
SMP	Software Management Plan
URL	Uniform Resource Locator
Walnuts	Walnuts Smart Snack Dietary Intervention Trial
WP	Work Package
WPL	Work Package Leader

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

1. Introduction

1.1 Project Overview

Empowering children and adolescents to adopt healthy lifestyles while tackling health inequalities is crucial. Person-centred interventions have been developed to this end but face several challenges: they are often restricted to one specific actor of children's environment (teachers, parents, or peers) and rarely target several significant others simultaneously, they are often restricted to one specific behavior (e.g., physical activity) or life domain (e.g., school), their effectiveness may differ according to socio-economic or socio-spatial conditions, and they are rarely both large-scale and personalised to individuals' needs. To address these challenges, the ambition of YEAH! is to develop an **innovative personalised person-centred digital intervention** that: (1) promotes children (8-10 years) and adolescents' (14-16 years) **social interactions** with multiple actors of their environment, (2) targets multiple **health behaviors** (physical activity, sedentary behaviors, diet, sleep), (3) combines in a novel manner evidence-based knowledge of **optimal intervention features** (theory-driven approach), with co-creation with end-users and stakeholders (data-driven approach) to tailor the intervention to the local context and target group, and (4) relies on a state-of-the-art method of intervention development to rigorously measure and verify its impact, optimise its **sustainable implementation** in the community and its **equity**, and evaluate its cost-effectiveness. To do so, YEAH!, will articulate interdisciplinary methods and concepts from psychology (health, social, developmental), public health, psychiatry, social marketing, geography, geomatics, environmental epidemiology, IT and computer sciences, statistics, physiology, and medicine, and on the participation of stakeholders (children and adolescents, parents, teachers, healthcare professionals, educators, local and national authorities).

1.2 Purpose of the deliverable

This deliverable provides the initial version of the DMP for the YEAH! project. Its primary purpose is to **guide the consortium** in the responsible **handling of research data** throughout the project lifecycle. It outlines how data will be **collected, generated, processed, documented, shared, curated, and preserved**, in alignment with the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles.

To ensure **Findability**, datasets will be described with rich metadata, assigned Persistent Identifiers (such as DOIs for datasets, ORCIDs for researchers, and SWIDs for software), and indexed in trusted repositories to facilitate their discovery. For **Accessibility**, data will be

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

made available through open and reliable platforms such as Zenodo. In terms of **Interoperability**, efforts will be made to adopt community-recognized metadata standards and controlled vocabularies (e.g., INRAE Thesaurus). **Reusability** will be promoted by providing clear licensing terms and thorough documentation. Preference will be given to open, non-proprietary formats (as outlined in Section 2.2), ensuring long-term access and facilitating re-use by third parties.

The DMP also describes the methodologies and standards to be applied, and clarifies which data will be made openly accessible and under what conditions.

For **project partners**, the DMP serves as a common reference to ensure coherent data practices across Work Packages (WPs), helping to manage and monitor data flows, responsibilities, and compliance with Horizon Europe (HE) requirements. It provides a project-wide overview of data types and flows, and facilitates collaboration by clarifying data-related expectations and processes.

For **external stakeholders**, including the European Commission (EC) and the broader scientific community, the DMP offers transparency on how the project's data is managed. It explains how data can be accessed, under which licenses, and how it may be reused, ensuring that knowledge generated in YEAH! contributes to open science and can be leveraged for future research, innovation, and societal benefit.

This initial DMP is aligned with the intellectual property rights (IPR) policy defined in the Consortium Agreement (CA) and will be updated throughout the project to reflect any changes in policies, practices, or data management needs.

2. Data Summary

2.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

WP	Purpose of the data that will be collected/generated for the project
WP1	The data collection begins with comprehensive systematic reviews to gather and synthesize existing knowledge on determinants of child and adolescent health behaviors and health. These reviews provide a foundation by identifying key factors and gaps, guiding the focus of subsequent empirical analyses.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Purpose of the data that will be collected/generated for the project
	The results obtained by the analyses on these data will be used to design the YEAH! intervention. Identified determinants of child health behaviors and health (obtained from the data-driven exposome analyses), and an understanding of when certain determinants are important and for which subgroup of the population (obtained from the hypothesis-driven analyses) will serve as targets for the intervention and will help to tailor the intervention. In addition, the identified socio-environmental determinants will inform the creation of vulnerability maps indicating the neighborhoods with highest vulnerability for low adherence to healthy behaviors and low well-being and health in children and adolescents. These vulnerability maps will be used to ensure sufficient recruitment of the most vulnerable participants in the large multicenter Randomized Controlled Trial (7 European countries) evaluating the YEAH! intervention. Finally, the results from the reviews and data-analyses on the knowledge base of determinants of child and adolescent health behaviors and health will be brought together in an online platform to (i) guide the development of the intervention content, (ii) guide the analysis of the heterogeneity of the RCT results due to contextual differences and (iii) disseminate the results to relevant stakeholders and the general public.
WP2	Data are collected as part of the co-creation process to enable the co-creator to understand a problem and design an appropriate solution
WP3	A web and mobile application will be developed as part of this Work Package (WP). This application will have several functions. First, it will be used to deliver educational material to the participants. It will also collect information on the level of physical activity performed by the participants, as well as their sleep, screen time and dietary habits, either through data from an activity tracker or by allowing participants to complete questionnaires. Finally, it will also serve as a space for interaction between healthcare professionals, peer leaders, children, and their parents.
WP4	The aim of this WP is to select the tools to be used in the evaluation of the YEAH! Intervention (WP8) and to validate the assessment instruments used for children

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Purpose of the data that will be collected/generated for the project
	and adolescents in the 6 languages in which the Randomized Controlled Trial (RCT) is implemented (French, Dutch, Danish, Romanian, Greek, and English).
WP5	The WP5 is the conception of the Randomized Controlled Trial that will be implemented in the WP6 and evaluated in the WP8. The data collected in the WP5 will be the data generated by WP2 and WP3. These data will serve to design the RCT.
WP6	The WP6 consists in the implementation of the Randomized Controlled Trial in the 7 participating countries to evaluate the YEAH! Intervention. The data collected in the WP6 will be collected based on the tools selected in WP4, and will serve to evaluate the outcome of the WP8: - Evaluation of the effectiveness of the intervention and mechanisms of action. - Evaluation of the equity of the intervention. - Evaluation of the external validity of the intervention. - Evaluation of the cost-effectiveness of the intervention. - Evaluation of the reach, adoption, implementation, and maintenance of the intervention. The data will be collected in at least 384 pupils (from 2 classes for each of the 8 schools) in each country, for a total of 2688 pupils.
WP7	The data collected in the YEAH! N-of-1 study will serve to evaluate the effectiveness of a holistic behavioral intervention targeting multiple health behaviors (physical activity, sedentary behaviors, diet, sleep) and actors of adolescents' environments (coaches, healthcare professionals). The objective is to assess short-term changes in health behaviors, well-being, and weight status among vulnerable adolescents—student-athletes and youths with mental disorders—through continuous monitoring and qualitative feedback. Data will also help identify mechanisms of action and contextual factors influencing intervention effectiveness, guiding future adaptations and implementation.
WP8	WP8 aims to evaluate the interventions in terms of effectiveness and equity and is therefore among the primary goals of trial data collection. The goal is to compare pre- and post-intervention metrics to track behavioral and systemic

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Purpose of the data that will be collected/generated for the project
	changes. This should support decision-making for policymakers and stakeholders regarding intervention scalability and equity.
WP9 – WP10	To monitor and evaluate communication and dissemination efforts, partners involved in WP9 and WP10 will collect data on outreach activities, stakeholder engagement, and visibility metrics. This contributes to the project's broader objectives of ensuring the uptake, accessibility, and long-term impact of the project's results.
WP11 – WP12	Data will be collected to ensure the effective governance and progress tracking of the project. This includes data on project planning, reporting, decision-making processes, consortium communication, and compliance with funding requirements. These data support the smooth implementation of the project and ensure that objectives are met on time and within budget.

2.2 What types and formats of data will the project generate or collect?

WP	Type	Format
WP1	Primarily numeric data, organized in databases and spreadsheets, as well as codebooks in spreadsheet format, separately for each of the four cohorts (ABCD, BiB, BREATHE & WALNUTS), with a clear folder hierarchy. The data will include information on determinants, outcomes, and confounders; and will use persistent identifiers (PIDs).	Data will be stored in open and widely used formats such as .csv and .xlsx.
WP2	The co-creation process generates mainly qualitative data from conversation and diverse collective intelligence and creative activities.	The data will come in multiple formats: audio recordings and transcription of conversations and focus groups, drawings and sketches, 3d images and models, other types of artefacts,

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Type	Format
		written notes, photographs and pictures. .mp3, .png, .jgeg, .docx, .pdf
WP3	1. Application specification 2. Application code 3. Behavioral and activity data 4. Personalized recommendation code 5. Intervention material 6. User data	1. .docx, .txt, .odt 2. programming language to be define based on the required features. 3. .raw, .csv, .docx, .pdf, .txt, .odt 4. programming language to be define in line with application code language 5. jpg, .png, .svg, .mp4, .mov, .avi, .pdf, .txt, .csv,.docx, .pptx, .odt 6. to be define based on required information
WP4	Translation/adaptation documentation, psychometric response data, metadata/codebooks, statistical outputs, consent and ethics documentation	.docx, .pdf, .xlsx, .csv, .sav, .json, .r, .rmd, .png
WP5	Data generated will include project reports, meeting minutes, deliverables, presentations. It will also include documents such as trial protocol, case report form, tutorial, training course, communication tools and various internal documents.	.pdf, .csv, .pptx, .odt, .docx, .xlsx
WP6	Numeric data: - Time spent in moderate-to-vigorous physical activity - Time spent in sedentary behaviors in general and in screen time in particular - Quality of diet	pdf, .csv, .pptx, .odt, .docx, .xlsx

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Type	Format
	- Duration and quality of sleep - Health behaviors - Well-being: subjective well-being, that reflects high life satisfaction, high positive affect, and low negative affect and social well-being, which includes social integration, social acceptance, social contribution, social actualization, and social coherence - Weight status - Sociodemographic data of the participants - Social network indices of the participants - Health consumption - Quality of life Data generated will include project reports, meeting minutes, deliverables, presentations, tracking tables and correspondence.	
WP7	Numeric data: time-series from accelerometers; survey responses on diet, well-being, and psychological constructs, collected via REDCap; weight and anthropometric data Textual data: transcripts of semi-structured interviews and focus groups; internal documentation; reports and presentations Audio recording Spatiotemporal data	.csv .docx, .pdf, .pptx .mp3 .rdf or ttl
WP8	While no further data should be collected/ generated from in this part of the project during the analysis derivatives of supplied	Depending on supplied data formats by previous WPs any or all of for example: .pdf, .txt, .csv, .xlsx, .dta

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Type	Format
	trial datasets from earlier WPs can be prepared.	
WP9 – WP10	Data collection on the Communication & Dissemination (C&D) activities carried out in the project including tracking of communication outputs, stakeholder engagement metrics, online platform analytics (website, LinkedIn), event participation data, and clustering activities with other projects.	.jpg, .png, .svg, .mp4, .mov, .avi, .pdf, .txt, .csv, .pptx, .odt, .docx, .xlsx
WP11 – WP12	Data generated will include project reports, meeting minutes, deliverables, presentations, financial and administrative records. Additionally, some data might be generated from project management tools (e.g., Gantt charts, task trackers) and collaborative platforms.	.pdf, .csv, .pptx, .odt, .docx, .xlsx

2.3 How will new data be collected or produced?

WP	Methodology/Software	Documentation of data provenance
WP1	During systematic reviews and/or meta analyses, relevant information will be extracted from published studies. Data extraction will use specialized review software (e.g., Covidence). New GIS/geocode data will be collected by sourcing publicly available or institutional geospatial datasets (e.g., satellite imagery, administrative boundaries, environmental measures).	Extracted data from previous research will be stored in .csv and .xlsx files. The geocode data will be saved in .csv files.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Methodology/Software	Documentation of data provenance
WP2	New data is collected throughout the co-creation process (usually 8 sessions of 2-4 hours). This will use a variety of methods ranging from brainstorming to system mapping. These methods are to enable people to work together to design solutions and be collectively creative and cooperate. It is unlikely that we will use specific software but we might use platforms such as Kumu or Mural which are online	Each session will adhere to a strict protocol that is defined by a cocreation process protocol. All data will be documented for each session and labelled with the session code and activity code.
WP3	App specifications will be defined in consultation with the members of the WP, in connection with the work carried out in WP2. This work will be conducted using collaborative editing tools such as Resana (the collaborative suite used by the French public sector). Application and Personalized recommendation algorithm code will be produced using a development environment such as VSCode. Documentation will be generated using tools such as doxygen. The overall source code will be stored in a versioning tool (git) such as gitlab. Behavioral and activity data will be collected either using an activity tracker (which will be defined in the near future) or by the application using tools such as forms. The production of the educational material will be aligned with the directions set out in WP2. Consequently, the methodology to be implemented will depend heavily on these	Application specifications data will be documented through periodic internal reports as well as a final report. Behavioral and activity data will be documented in a readme file indicating a short description, the type of data (continuous, discrete, ...) and associated units. User data will be documented in a readme file indicating a short description, the type of data and associated units.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Methodology/Software	Documentation of data provenance
	outcomes and will be defined in a later version. User data will be created and collected throw the application using registration forms.	
WP4	Translation and adaptation follow International Test Commission guidelines; questionnaires administered via REDCap or Qualtrics or other software; statistical analysis conducted using R/RStudio.	Version-controlled translation/adaptation files; participant data tagged with anonymised IDs; metadata and analysis scripts stored in shared repository (e.g. OSF, Zenodo).
WP5	New data will be produced through internal reporting processes, meeting documentation and correspondence.	Archiving and versioning of the document produced.
WP6	Accelerometers (to be specified once the devices are validated) for continuous tracking of physical activity, sedentary behavior, and sleep. Online questionnaires via REDCap for diet, screen time, sleep, health behaviors, well-being, weight status, socio-demographic data, social network indices, health consumption and quality of life	Raw data timestamped and labelled by participant ID; metadata include device type, placement, calibration logs; data stored in .csv format with README file detailing variables. Each survey entry is linked to participant ID; metadata documented in a codebook (.to be specified once the devices are validated) or .pdf).
WP7	Accelerometers (to be specified once the devices are validated) for continuous tracking of physical activity, sedentary behavior, and sleep	Raw data timestamped and labelled by participant ID; metadata include device type, placement, calibration logs; data

Funded by the European Union
This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Methodology/Software	Documentation of data provenance
	Online questionnaires via REDCap for diet, well-being, and psychological constructs Semi-structured interviews and focus groups (recorded via voice recorder) SaLTo model (Semantic Life Trajectories) for spatiotemporal data modelling Transcription software (NVivo) and qualitative coding tools	stored in .csv format with README file detailing variables. Each survey entry is linked to participant ID; metadata documented in a codebook (.to be specified once the devices are validated) or .pdf). Audio files linked to participant ID; metadata include interviewer, date, context; stored with README and consent forms. Each episode/event tagged with unique ID, source, and contextual metadata (location, date, source file); structured in RDF/TTL format; metadata file describing schema and vocabularies. Each transcript linked to original audio file and interview protocol; coding trace documented with memos and versioning; metadata include coder ID and date of transcription.
WP8	No new data is planned to be collected any derivatives will be documented in the logs and documentation of the data analysis.	Data pipelines, who analyzes and accesses what and when, variables, their transformations and usage in respective logfiles as well as versions of the analysis code and datasets using for example Git will be documentation methods

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Methodology/Software	Documentation of data provenance
WP9 – WP10	Collection of web analytics using tools such as Matomo or Google Analytics; social media metrics from LinkedIn; communication procedure feedback forms; internal progress report template; tracking of publication and media mentions	Data will be documented through periodic internal reports, analytics dashboards, event reports by the Project Management Team.
WP11 – WP12	New data will be produced through internal reporting processes, meeting documentation, budget tracking, and correspondence.	Archiving and versioning of the document produced.

2.4 Will you re-use any existing data and what will you re-use it for?

WP	Existing data that will be re-used
WP1	Existing data from four well-established cohorts (ABCD, BiB, BREATHE & WALNUTS) will be re-used. These cohort datasets provide comprehensive information on child and adolescent health behaviors, outcomes, and relevant determinants, making them highly valuable for our analyses while the project resources do not allow for the collection of these data. All data re-use will strictly adhere to the original consent agreements included in the Data Transfer Agreements (DTAs) and any applicable ethical and legal constraints.
WP2	Not applicable.
WP3	No specific data will be reused.
WP4	Previously validated measures (if available) will be used as the foundation for translation and adaptation into six languages. These existing instruments, identified from the Clinical Studies Annex of the Part B of YEAH! project (Table 1a and 1b), will be culturally adapted and validated in new populations (children and adolescents) to ensure reliability, construct validity, and measurement invariance across countries.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Existing data that will be re-used
WP5	Re-use of existing questionnaires translated and validated in the WP4. These questionnaires will be included in the design of the RCT. The best practices of the WP4 from which the data originates will be respected.
WP6	Re-use of existing questionnaires translated and validated in the WP4. These questionnaires will be used to collect the parameters defined in the protocol in order to evaluate the YEAH! Intervention. The best practices of the WP4 from which the data originates will be respected.
WP7	No existing datasets will be directly re-used in the YEAH! N-of-1 study. All data will be newly collected to address the specific research questions, which require intensive, individualised, and context-sensitive longitudinal monitoring of health behaviors and well-being in adolescents. Existing datasets were considered (e.g., previous studies on adolescent health behaviors), but were discarded due to incompatibility with the study design, absence of fine-grained temporal resolution, and lack of contextual and spatiotemporal information necessary for the N-of-1 approach and SaLTo modelling. Moreover, the intervention being co-developed and contextually adapted, only newly collected data can accurately capture its implementation and outcomes. No constraints regarding the re-use of external datasets apply to this study.
WP8	No specific research data will be reused or generated. The best practices of the WP6 from which the data originates will be respected.
WP9 – WP10	No specific data will be reused. UGA (WP Leader) will utilize internal data management tools and predefined templates to monitor and assess the impact of communication, dissemination, and networking activities throughout the project.
WP11 – WP12	No specific data will be reused.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

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

WP	Origin of the data that will be generated or re-used
WP1	Re-used Data: The primary datasets originate from four established cohorts that were previously collected as part of large-scale research studies on child and adolescent health. The provenance of these datasets is maintained through their original data custodians, ensuring traceability and authenticity. Newly Generated Data: New data will be generated through systematic reviews and meta analyses, where information is systematically extracted from published scientific studies. Additional new data will be collected from trusted, publicly available geospatial datasets (e.g., satellite imagery, administrative boundaries, environmental measures).
WP2	The data will be generated during the co-creation sessions only. We will not be re-using data because this has no purpose.
WP3	Behavioral as well as activity data will be collected by the application from 2688 participants (8-10y children and 14-16y adolescents) coming from 7 countries in Europe (France, Belgium, Netherlands, Denmark, Romania, Greece, United Kingdom). User data will be collected from any user of the application (teacher, parents, peer leader, children and adolescents, …) Origin and provenance of the intervention materials depend on his creation, it will then be defined in a later version.
WP4	Re-used data originate from existing validated psychological instruments listed in the Clinical Studies Annex of the Part B of YEAH! project (Table 1a and 1b), developed and published in peer-reviewed sources. New data will be generated through primary data collection from children (8–10 years) and adolescents (14–16 years) across participating countries via structured questionnaires. All data will be documented, anonymised, and version-controlled to ensure traceability.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Origin of the data that will be generated or re-used
WP5	Re-used data originate from existing validated psychological instruments listed in the Clinical Studies Annex of the Part B of YEAH! project (Table 1a and 1b), developed and published in peer-reviewed sources. The data generated originate from internal project activities, partners meetings, correspondence.
WP6	All data used will be newly generated for the purpose of this research. The data will originate directly from: Participants through self-reports and sensor-based monitoring. Members of their social environment, including coaches, and healthcare professionals, who will contribute via interviews and training feedback; Digital devices and software, such as accelerometers (for physical activity, sedentary behavior, and sleep) and online platforms like REDCap (for questionnaire data collection). Each dataset will be documented with metadata indicating the collection method, date, instrument used, and responsible researcher. Data provenance will be assured through traceability procedures, timestamping, and pseudonymised coding of participant data. No pre-existing datasets will be reused.
WP7	All data used in the YEAH! N-of-1 study will be newly generated for the purpose of this research. The data will originate directly from: Participants, i.e., adolescents aged 14–16 with poor mental health or enrolled in sport-study programmes, through self-reports, interviews, and sensor-based monitoring; Members of their social environment, including coaches, and healthcare professionals, who will contribute via interviews and training feedback; Digital devices and software, such as accelerometers (for physical activity, sedentary behavior, and sleep) and online platforms like REDCap (for questionnaire data collection); Each dataset will be documented with metadata indicating the collection method, date, instrument used, and responsible researcher. Data provenance will be assured through traceability procedures, timestamping, and pseudonymised coding of participant data. No pre-existing datasets will be reused.

Funded by the European Union
This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Origin of the data that will be generated or re-used
WP8	The origin would be the other Work Packages in particular WP6 and WP7.
WP9 – WP10	Data will be generated by monitoring the outcomes of dissemination and communication efforts, using desktop research and engagement in clustering activities with other related projects.
WP10 – WP11	The data generated originate from internal project activities, consortium meetings, and reports. Some data, such as the Grant Agreement, are provided by the funding agency and the project consortium.

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

WP	Data	Size of the data that will be generated or re-used
WP1	Datasets for the four cohorts, for the childhood period (ABCD, BiB, BREATHE) and the adolescent period (ABCD, BiB, WALNUTS).	The total size of the original data for one cohort is 250 MB. Including the enriched data, this becomes 2.5 GB per cohort, estimating total size at 15 GB.
WP2	Audio recording and images	Size of the data will be small. We expect less that 1 Gb by site and there are 7 sites. In total, up to 7 Gb.
WP3	1. Application specification 2. Application code 3. Behavioral and activity data 4. Personalized recommendation code 5. Intervention material 6. User data	1. < 100Mo 2. <2 Go 3. Between 100Go and 200Go 4. <500 Mo 5. Depend on the nature of data, to be specified in a future version. 6. <100go
WP4	Translation/adaptation documents, questionnaire responses from	Data size may vary slightly depending on the number of items per scale and the inclusion of additional validation

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data	Size of the data that will be generated or re-used
	children and adolescents, metadata, and statistical outputs	variables (e.g., physiological data, health behaviors, well-being). Depending on sample size, measurement frequency, and measurement tools up to 50GB.
WP5	Protocol, Case Report Form, project reports, meeting minutes, deliverables, presentations, tracking tables and correspondence.	2 to 5 Go
WP6	Accelerometers (to be specified once the devices are validated) for continuous tracking of physical activity, sedentary behavior, and sleep. Online questionnaires via REDCap for diet, screen time, sleep, health behaviours, well-being, well-being, weight status, socio-demographic data, social network indices, health consumption and quality of life	To be determined on the basis of the questionnaires selected from WP4 (not done yet). Approx. 2 to 5 Go
WP7	Accelerometer data (physical activity, sedentary behavior, sleep) Online questionnaire data (diet, well-being, psychological constructs) Transcripts of interviews and focus groups (text files) Audio recordings of interviews (if consent obtained)	~10–15 GB (20 student-athletes + 24 adolescents × continuous recording over 12 weeks) ~2–3 MB (exported in .csv format) ~5–8 MB (1 interview per participant and 1 focus group per stakeholder group) ~8–12 GB (compressed .mp3 format; average 30–60 min per interview)

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data	Size of the data that will be generated or re-used
	SaLTo semantic life trajectory data (student-athlete sub-sample) Metadata and documentation files (README, codebooks, license, author list)	~1–2 GB (RDF or TTL format for ~20 individuals) ~5–10 MB **Total estimated size**: ~20–30 GB (depending on consent for audio recording and completeness of data collection)
WP8	Trial Data passed on from other WPs, the as well as any transformations of such to ensure high quality analysis	The sum of trial data gathered and cleaned for analysis and evaluation purposes. Depending on sample size, measurement frequency, and measurement tools up to 50GB.
WP9 – WP10	Data related to the tracking of dissemination and communication activities, including reports, presentations, metrics, and impact data.	Between 200 Mo et 2 Go
WP10 – WP11	Document files, spreadsheets, and presentations.	~2 Go

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

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

Sector	Stakeholders
Projects/Initiatives	Other EU-funded research projects, especially those focused on public health, social sciences, or digital health solutions. The data can support collaborative efforts and provide benchmarks for effective dissemination practices.
Industry	Companies and organizations involved in the development, production, and distribution of digital health technologies, wellness products, and related services. Data can assist in the development of new products, digital interventions, and services aimed at improving health among young populations.
Public Health and Education	Educational institutions, public health authorities, schools, and child health organizations. Data could inform strategies to promote healthy lifestyles in educational settings and public health campaigns.
Policy Makers and Government	EU bodies, national ministries of health, and local government departments. Data will support evidence-based policy decisions on health interventions for children and adolescents.
Scientific and Research Community	Research and Technology Organizations, researchers, academics, experts and institutions involved in health promotion, digital health, and related fields. Data will help advance knowledge in health behavior, lifestyle choices, and preventive health.
NGOs and Civil Society	Non-governmental organizations (NGOs), community groups, and advocacy organizations focused on child health, education, and well-being. Data can help shape programs for healthier lifestyle promotion.
General Public	Civil society, citizens of all ages, gender and socio-economic status and/or background, local communities and civil society organizations. Data can empower communities by raising

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Sector	Stakeholders
	awareness about healthy lifestyle choices and the importance of health promotion for children and adolescents.
International Health Organizations	WHO, UNICEF, and other global health organizations. Data may inform international efforts to promote child and adolescent health, particularly in diverse socio-economic contexts.

2.8 How and with what resources will the data be processed/analyzed?

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
WP1	All re-used cohort data will be harmonized following a harmonization protocol to align variables, coding schemes, and measurement units across the four cohorts. Analyses will be conducted following pre-specified detailed analysis protocols (Hypothesis-driven analysis protocol & Exposome analysis protocol). For meta-analysis and review research, all data will be extracted in line with a pre-registered review of review protocol, and follow analysis protocols specified as per the review type.	Statistical Software: - R (including packages for meta-analysis, exposome, and epidemiological analysis) - SPSS - Comprehensive Meta-Analysis (CMA) Data Storage and Management: Data will be stored and managed on secure institutional servers and/or trusted repositories that ensure compliance with FAIR data principles.	Research staff hired on the YEAH! project pertaining to WP1 partners involved in analyses tasks, including four institutions: ACCARE, ENS, IISPV and ISGlobal.	Statistical Software: R is free; for SPSS/CMA an already existing license will be used. Servers/Repositories: institutional or costs covered by each institution. The purchase or use of desktops, laptops, or workstations equipped with the necessary hardware and software to handle data storage, processing, and analysis tasks, are covered under the WP1 budget. No specific additional analysis costs are foreseen.
WP2	The only processing that data will undergo is transcription, translation to English and anonymisation	We will use aiforcocreation platform for automatic translation and transcription	WP 2 staff and local facilitator will read and analyse the data to plan subsequent co-creation sessions and evaluate the co-creation process. Co-creators will	The main form of analysis will be concept mapping and topic maps generation. No use paid software, no storage costs and

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
			have access to all the data to refer to during the cocreation. Staff involved is internal to the project team and holds skills in co-creation qualitative analysis.	no costs for a computing server.
WP3	Behavioral and activity data will undergo a cleaning process to ensure their validity and to eliminate any duplicates. They will then will be processed in order to create an algorithm capable of providing recommendations of behavior changes to the participants. User data will be anonymized in order to comply with the GDPR. Behavioral and activity data will also be used to quantify changes and, consequently, the effectiveness of the Yeah! Project (done in WP8).	Cleaning and duplicates deletion as well as anonymization will be done by the application. Processing and algorithm development tool will be defined in a future version.	WP3 members Researchers and engineer from UR2 will access the whole content. Researcher and post-doc from ENS will access Behavioral and activity data Researchers, post-doc, engineers from WP8 members will access Behavioral and activity data Every researcher is internal to the project and engineers and post-doc are specifically hired for this purpose.	Covered under WP3 budget; no additional specific costs foreseen. No use paid software, no storage costs and no costs for a computing server.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
WP4	Data cleaning, scoring, and psychometric validation (e.g., internal consistency, factor analysis, measurement invariance) following pre-defined statistical analysis plans. Translation/adaptation follows ITC guidelines.	R/RStudio for statistical analysis (R version 4.4.1 (2024-06-14); RStudion version 2023.9.1.494); REDCap or Qualtrics or other software for data collection; institutional servers and cloud-based platforms (e.g., OSF, OneDrive) for storage and sharing. The choice between REDCap and Qualtrics for data collection has not been made yet – probably REDCap, as the intervention will use REDCap. REDCap is free, Qualtrics is paid.	Research staff at UBB and partner institutions involved in WP4; data analysts, and PhD/postdoctoral researchers. We will recruit a postdoc specifically for data analysis with skills in advances statistics knowledge and analyzing large databases	Covered under personnel costs in the project budget; no major additional costs expected for software (open-source) or infrastructure (institutional support). No use paid software, no storage costs and no costs for a computing server.
WP5	Data will be processed and managed using standard office software (e.g., Microsoft Office suite), and communication platforms (e.g., email, videoconference).	Standard office equipment (laptops), secured servers, collaborative platform (ShareDocs)	Internal project researchers Project staff at UGA and partners sites involved in the project Behaviors training coaches with specific skills will be recruited specifically for the project	Covered under WP5 budget; no additional specific costs foreseen since already paid by UGA: - licenses for IT tools (Microsoft Office suite, tools for email, videoconferences, anti-

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
				virus and anti-phishing softwares) - storage costs and computing server
WP6	Some data will be collected from accelerometer. The online questionnaires will be filled by the participants using the REDCap platform. RCT management data will be processed and managed using standard office software (e.g., Microsoft Office suite), and communication.	REDCap (server software, no need to install on a computer, online use) Standard office equipment (laptops), secured servers, collaborative platform (ShareDocs).	Internal project researchers Project staff at UGA and partners sites involved in the project Behaviors training coaches with specific skills will be recruited specifically for the project	REDCap is provided and hosted by UGA. There is no additional cost for the collection of the data of the project. Covered under WP6 budget; no additional specific costs foreseen.
WP7	Cleaning and analysis of accelerometer and questionnaire data, transcription and thematic coding of interviews, modelling of life	R, NVivo, REDCap, Office/R Markdown will be used, with the latest available versions at the time of implementation. Since WP7	Biostatisticians, qualitative researchers, data scientists and project staff at UGA and partner sites - include both internal	We will require a license for NVivo for qualitative data analysis. All other software tools

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
	trajectories (SaLTo), and integration of mixed data	will start at a later stage, the specific versions will be detailed in the next update of the DMP.	personnel from UGA and external collaborators from partner sites. Some are already affiliated with the institutions involved, while a post-doc will be specifically recruited for tis WP.	used in the project will be open-source or freely available. Data storage will be managed using the infrastructure provided by the partner institutions, without generating additional costs. The WP7 budget includes funding for personnel (notably a post-doctoral researcher), the NVivo license, and minor operational expenses, such as travel for data collection and potentially the purchase of specific equipment (e.g., activity tracers).
WP8	Data lineage tracking, Version-control of dataset and scripts Cleaning and analysis of the health consumption and quality of life data	Readmes of version control and the data pipeline as well as a master version of the processing and analysis steps should give an overview of the data flow	Research staff hired at UHEI for YEAH! including Senior and junior researchers with experience in quantitave data analysis PI at	No storage costs or computing costs are anticipated; For paid software is to be determined whether analysis will be run using Stata (annual

Funded by the European Union
This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
		Softwarewise, depending on how the team finally carrying out the analyses will be composed and what is planned to be done specifically it will be R or STATA as analysis software. Stata 19 or later;	UHEI. It will be limited to project partners only. Medico-economic evaluation team	prescription around 700$) for two years
WP9 – WP10	Quantitative and qualitative tracking of engagement (e.g website and LinkedIn visits, downloads, event attendance, social media analytics…); analysis of clustering interactions	Excel files and internal data tracking templates, analytics platforms (Google Analytics, Matomo) social media insights tools. For the moment, we continue using Google Analytics due to its powerful search engine integration and widespread adoption, which allows for more comprehensive benchmarking and traffic analysis. However, we are transitioning gradually to Matomo, an open-source and privacy-compliant alternative, to enhance data	Coordinator, Project Manager (recruited for the project), Communications Officer (recruited for the project) and the whole consortium.	No use paid software, no storage costs and no costs for a computing server.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

WP	Data Processing Methods/Protocols	Equipment/Technical Platform	Research Staff	Data Treatment/Analysis Costs
		protection and align more closely with GDPR principles.		
WP10 – WP11	Data will be processed and managed using standard office software (e.g., Microsoft Office suite), and communication platforms (e.g., email, videoconference).	Standard office equipment (laptops), secured servers, collaborative platform (ShareDocs)	Coordinator, Project Manager (recruited for the project), and the whole consortium.	No use paid software, and no costs for a computing server. For storage, SUMMER will be used, and a budget will be necessary (to be determined later).

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

3. Data Management Policy: FAIR data

According to the OpenAIRE definition, the four basics of FAIR are:

- **Findable**: discoverable with metadata, identifiable and locatable by means of a standard identification mechanism
- **Accessible**: always available and obtainable; even if the data is restricted, the metadata is open
- **Interoperable**: both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organisations or countries
- **Reusable**: sufficiently described and shared with the least restrictive licenses, allowing the widest reuse possible and the least cumbersome integration with other data sources.

3.1 Making data findable, including provisions for metadata

DOI

All open data, publications, and open-source software generated through the YEAH! project will be made easily findable via persistent **Uniform Resource Locators** (URLs). Whenever possible, a **Digital Object Identifier** (DOI) will also be assigned to ensure unique and reliable citation of these outputs.

Materials shared through **Zenodo**, the project's default open-access repository, will automatically receive a DOI, with the added benefit of Zenodo's built-in versioning support.

In cases where outputs are hosted in institutional repositories, scientific publisher platforms, or other research data repositories, they will at minimum be accessible through a **permanent URL**. If the hosting institution is affiliated with a DOI registration agency, a DOI may also be issued.

The type of unique identifier assigned to scientific publications (such as a DOI, Publisher Item Identifier [PII], or International Standard Serial Number [ISSN]) will depend on the chosen open-access route (green or gold), the editorial policy of the journal, and the repository used.

Naming conventions

Files and folders stored in data repositories will follow a **standardized naming convention** that includes **versioning** and ensures a clear, consistent structure.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

[YEAH!]_[Date]_[Partner Abbreviation]_[WP#]_[Document Type]_[Confidential]_[Document Title]_[Version]

In blue, the optional information. Each component corresponds to:

- YEAH!: The name of the project
- Date: The creation or modification date in the format YYYYMMDD
- Partner Abbreviation (optional): The abbreviation of the author partner name
- WP# (optional): The Work Package number
- Document Type: Type of the document
- Confidential (optional): only if the document is confidential
- Document Title: A short description of the document or its title.
- Version: The version number and the state of the document (Draft, Final, Approved)

Further details can be found in the *Project Management Plan, Section 4.3 File Naming Conventions*.

Zenodo

Open data generated by the YEAH! project will be deposited in **Zenodo**, an open-access digital repository that supports research data sharing in accordance with the **FAIR principles** (Findable, Accessible, Interoperable, Reusable). Developed by **OpenAIRE** and hosted by **CERN Data Centre**, Zenodo is designed to support a wide range of research outputs and facilitate their discoverability and reuse.

Zenodo allows researchers to **upload various types of content**, such as publications, datasets, images, videos, and more, across all scientific disciplines. It also provides **persistent identifiers** like DOIs, ensuring that each deposit is uniquely citable and traceable. This makes it easier for researchers to showcase their work, receive proper attribution, and comply with open science requirements set by funders like the EC.

By using Zenodo, YEAH! ensures that all project results can be easily shared, accessed, and reused, while being properly integrated into the **OpenAIRE portal** and linked to relevant research outputs. A dedicated Zenodo community has been set up for YEAH!, to gather and showcase all related outputs in one centralized space.

Search keywords

Zenodo allows to perform simple and advanced search queries on Zenodo using the keywords (see user guide). Project team members responsible for the datasets will be tasked with uploading the public datasets they have generated and assigning **specific**

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

keywords that describe the content of each dataset. It is important that these specific keywords accurately reflect the dataset's content.

Moreover, to project-specific terms, it is recommended to use controlled vocabularies to ensure consistency and interoperability of metadata. For this purpose, the INRAE Thesaurus can be consulted. This open thesaurus provides a structured vocabulary designed to support metadata standardization.

In addition, it is requested that the **general keywords from YEAH!** project must be included in all public datasets, scientific publications, and public deliverables: behavioral science, health behavior change, person-centered intervention, health, children, adolescents, health inequalities, empowerment, healthy lifestyle, personalised approach.

Specific additional keywords for each dataset may be added as appropriate by the data owners.

Version numbers

Individual file names, datasets, and publications will contain version numbers that will be incremented at each revision (Vxyz).

Zenodo offers **DOI versioning for all datasets** shared within a community. This feature enables us to edit and update datasets even after they have been published. Additionally, it allows us to cite specific versions or all versions of a dataset.

Metadata

Metadata refers to the **descriptive information** that accompanies research data and makes it easier to **find, understand, and reuse**. It is crucial for ensuring that datasets can be easily found in online repositories.

By including comprehensive and detailed metadata, other researchers can quickly assess the relevance and potential value of a dataset for their own work. To ensure interoperability and adherence to the FAIR principles, metadata should ideally be provided in a standardized format. This means using a common structure and vocabulary that allows machines and humans to interpret the metadata consistently across different platforms.

For now, **the specific metadata standard to be used in YEAH! project has not been defined**. However, it will be selected at a later stage to ensure compatibility with the chosen data repository (Zenodo) and to support FAIR data principles.

We will give preference to the **Dublin Core** metadata standard, which is widely supported and compatible with Zenodo. A useful guide can be found here.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Each WP may choose to adopt a different metadata standard if necessary, depending on the type of data produced. In such cases, the decision must be made at the start of the WP and justified accordingly. If no specific need arises, **Dublin Core will be the default standard**.

The metadata will be stored separately from the raw research data and kept updated throughout the project lifecycle.

In line with HE requirements, the following table summarises the minimum metadata elements expected for datasets and scientific publications to ensure compliance with FAIR principles and Open Access obligations.

Table 1: Comparison of Metadata Requirements for Publications and Datasets. CC0: Creative Commons Public Domain Dedication; FAIR: Findable, Accessible, Interoperable, Reusable; PID: Persistent Identifier.

Metadata Element	Data	Publications
Open license requirement	Creative Common Public Domain Dedication or equivalent (to the extent legitimate interests or constraints are safeguarded)	Creative Common Public Domain Dedication or equivalent
FAIR principles compliance	Yes (particularly machine-actionable)	Yes (particularly machine-actionable)
Information provided	Dataset descriptions, date of deposit, author(s), and embargo	Publication author(s), title, date of publication, publication venue
Funding information	Horizon Europe funding; grant project name, acronym, and grant number	Horizon Europe funding; grant project name, acronym, and grant number
Licensing terms	Must be included	Must be included
Persistent Identifiers (PIDs)	For dataset, authors, and if possible: their organisations and the grant	For publication, authors, and if possible their organisations and the grant
Related research outputs	If applicable: PIDs for related publications and other research outputs	If applicable: PIDs for any research output or tools/instruments needed to validate the conclusions of the publication

This basic metadata structure can be expanded with additional fields as needed, based on requests from the consortium.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

3.2 Making data accessible: Repository

Trusted repositories for data access

Datasets that are shared openly on trusted repositories will remain accessible for a minimum of 5 years. Each repository considered will be evaluated using a checklist to verify key aspects such as the **persistence of metadata** beyond the data retention period, repository **reliability**, and **compliance with the FAIR principles**.

The primary open-access repository identified for the YEAH! project is **Zenodo**, a trusted repository operated by CERN and supported by OpenAIRE. Zenodo is widely used in HE projects and does not require prior arrangements. It supports DOIs, rich metadata, versioning, and licensing options.

Project-specific secured repository

Sensitive data will be encrypted; for more details, please refer to section 4.4 *Sensitive data*. Given the volume of data collected and produced, SUMMER, a **specific ad-hoc secured central repository** for pseudonymised data and hosted in France (when in accordance with national laws and regulations) will be maintained by the Project Coordinator (UGA), and will remain operative 5 years after the project ends, except for cohort data (WP1) which will stay with their owners (in this case data will be shared only at the level of pooled results). Moreover, all the datasets will also be stored in each of the participant entities databases.

Assignment and resolution of persistent identifiers

Zenodo automatically assigns a **DOI** to each deposited dataset, ensuring it can be uniquely and persistently cited and accessed. These DOIs are resolvable to digital objects and are interoperable across repositories and indexing systems. Metadata and datasets will remain accessible even in the long term, supporting the reuse of results by the broader research community.

Repository coordination and planning

Although no formal agreement is required to use Zenodo, the Project Management Team (PMT) will assist partners in depositing data correctly, ensuring consistent metadata, licensing, and FAIR compliance. The checklist used for evaluating repositories will be kept by the Project Coordinator and used when alternative repositories are considered.

To ensure the broadest impact of YEAH! research, the project's results will be **shared both internally** within the consortium and **externally** with the wider scientific and stakeholder

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

communities. Data and results will be made available through **various channels**, including publications journals, conference presentations, and by being uploaded to open-access (OA) data repositories.

YEAH! results will adhere to the **OA requirements** outlined in the Annoted Grant Agreement, Ann*ex 5 - COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY (— ARTICLE 17) (HE)*, available in Annex VI, ensuring that research outputs are openly accessible to the scientific community and beyond.

"The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications
- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or a license with equivalent rights; for monographs and other long-text formats, the license may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and
- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements."

Open Data and Accessibility

The datasets will first be stored and organized by the respective data owners (either on personal systems or secure institutional servers) and in SUMMER, the ad-hoc central repository. These data will be made **accessible for verification and reuse**, unless a Work Package Leader (WPL) provides a valid reason to restrict access. To safeguard the project's intellectual property, **Creative Commons licenses** will be applied where appropriate.

The datasets will be publicly available, with no restrictions by default (if full anonymisation is possible; otherwise, they will be made available with restricted access) on at least one of the following platforms:

- OpenAIRE
- Zenodo

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

All datasets uploaded to Zenodo will be freely available to the public.

All open project outputs, including data, software, and scientific publications, will be made freely available through a suitable OA repository, such as Zenodo, as soon as feasible. Specifically, any research data required to validate findings in scientific publications will be uploaded to a data repository simultaneously with the publication itself. Research data that cannot be made publicly available will be stored in the repository with restricted access.

Scientific publications

To ensure OA for peer-reviewed scientific publications, YEAH! will publish in OA journals that follow the **green, diamond, or gold OA models**, with or without author processing fees. These publications, along with their bibliographic metadata, will be publicly available as OA. These will be announced on the project website and the R&I Participant Portal. To streamline reporting, publications should be deposited in an **OpenAIRE-compliant** repository, either by the authors (green OA) or by the publisher (diamond or gold OA). While OA papers can be shared via academic social networks like ResearchGate, the publication must first be deposited in a suitable OA repository.

Only publication fees in full open access venues for scientific publications are eligible for reimbursement. The project's preference is for **green OA** or diamond **OA journals** without author fees, but the journal's visibility, reputation, and publication speed will also be considered. In case of doubt, the OA requirements will be checked via the Journal Checker Tool.

- **Green OA (Self-Archiving):** Green OA refers to self-archiving, where the author archives the final peer-reviewed version of the article in an online repository, usually after publication in a journal. The journal must grant permission for self-archiving within 12 months of publication. To help identify journals that permit self-archiving and to understand the specific conditions (e.g., embargo duration, version allowed, licensing), researchers can consult the Search Open Policy Finder. This tool provides up-to-date information on publishers' open access policies and their alignment with funders' requirements, such as those of Horizon Europe.
- **Diamond and Gold OA (OA Publishing):** Diamond and Gold OA refer to publications made freely available by the publisher. Some gold OA journals require an author-processing fee, while others, known as diamond OA, do not. Fees for gold OA journals can be reimbursed within the project's budget. Some publishers allow authors to deposit a copy of the article in a repository, sometimes with an embargo period. Researchers can consult the Directory of OA Journals, a service that indexes high

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

quality, peer-reviewed OA academic journals that use an appropriate quality control system.

Accessing open data stored in a repository generally does not require any specific tools or advanced methods. Files can be easily downloaded using a standard internet browser. However, the way the data can be viewed, interpreted, processed, or modified offline will depend on its format and structure.

Data sharing exceptions

While the goal is to make all project data and publications openly accessible, several exceptions may apply:

- **Copyright and third-party data permissions**: When datasets are created by processing or combining data from multiple sources, the resulting outputs may involve complex intellectual property issues. These datasets can only be shared openly if all source data (e.g., input for models) is also openly accessible. In such cases, users must contact either the Intellectual Property Rights (IPR) team of their institutions or the original data owners to request access. If needed, legal agreements such as non-disclosure agreements (NDAs) may be put in place.
- **Confidentiality and personal data protection**: Datasets involving sensitive information, such as data about individuals or critical infrastructure, are not openly available by default, as sharing them may raise privacy, ethical, or security concerns. By definition, **WP2** will generate sensitive data that cannot be publicly shared. However, if full anonymisation of certain datasets is possible, these may be made available accordingly.
- **Commercial exploitation of data**: Data that is intended for commercial use and part of a business strategy will not be shared openly to protect its economic value.

Sensitive data restrictions

Each WP will indicate the list of datasets and specify the associated access in *Annex II: Datasets collected in the YEAH! Project*. For datasets that cannot be shared openly due to sensitivity, **controlled access** will be granted. In this case, the following procedure will apply:

- Each project partner limiting access to certain datasets will define **internal procedures** for identifying those authorized to access the data. These procedures will define roles, responsibilities, and criteria for granting access.
- The Project Coordinator has been designated as the Data Manager. Depending on the project's evolving needs, the review process may be delegated to another structure at a later stage.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

For datasets involving personal or sensitive information, partners should refer to their respective Data Protection Officers (DPOs) for guidance and compliance. Below is a table listing the DPO contact details for each project partner.

Table 2: DPO Contact Details per Project Partner

Partner Name	Data Protection Officer (DPO) contact
UGA	dpo@grenet.fr
USMB	relaisdpo@univ-smb.fr
UBB	dpo@ubbcluj.ro
SDU	dpo@sdu.dk
ISGLOBAL	lopd@isglobal.org
UTH	dpo@uth.gr
GCU	dataprotection@gcu.ac.uk
CHU CF	macousseau@chu-clermontferrand.fr
UHEI	datenschutz@uni-heidelberg.de
ULB	rgpd@ulb.be
JYU	riikka.h.valkonen@jyu.fi tietosuoja@jyu.fi
FLVS	stephane.sibalo@vivonsenforme.org
ENS Rennes	dpo@ens-rennes.fr
ACCARE	fg@accare.nl
IISPV	dpd@iispv.cat
RENNES 2	dpo@univ-rennes2.fr

In cases where open access cannot be granted due to sensitivity concerns, efforts will be made to share data under controlled access conditions. If such data is stored in **Zenodo**

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

with restricted or embargoed access, a **description of the dataset and metadata** will still be made publicly visible. Metadata will indicate when and under what conditions the data may become available.

Metadata for open, closed, embargoed, and restricted records **are always publicly accessible in Zenodo**. However, data files and datasets marked for restricted access are visible only to the owner and individuals granted permission by the owner. The restricted access option allows a researcher to upload a dataset while specifying the conditions under which access is granted. Researchers wishing to request access must justify how they meet these conditions. The dataset owner is notified of each new request and has the discretion to either approve or deny it. If approved, the requester is provided with a private link, which typically expires within 1 to 12 months.

3.3 Making data interoperable

General data and metadata vocabularies, standards, formats, and methodologies will be adhered to wherever possible. However, as no universally accepted field-specific vocabularies, standards, or formats currently exist, project partners will make efforts to apply widely recognized best practices in data interoperability. Before implementing these, it will be necessary to confirm that the majority of partners have the technical capacity to support them.

The project will need to develop its own project-specific ontologies and vocabularies. In doing so, the team will draw inspiration from existing ontologies and vocabularies to promote consistency and harmonization. Once a preliminary version of these resources is approved, they will be made openly available and updated on a regular basis. Increase data re-use

- All datasets that can be made openly available will be **shared under suitable open licenses** (e.g. Creative Commons), ensuring legal reuse, modification, and redistribution.
- **Metadata will be openly accessible** under a CC0 license, even for restricted or embargoed datasets, enabling discoverability and future controlled access.
- Research outputs will be deposited in trusted open repositories such as Zenodo, which assign **DOIs** and ensure long-term preservation and findability.
- All metadata will comply with **FAIR principles** and include information such as project name, grant number, licensing terms, and persistent identifiers (details in section 3.1)

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Data will be stored and shared using **standards and interoperable formats** to facilitate reuse across disciplines.
- For datasets not eligible for full open access (e.g. personal data, commercially sensitive data), a procedure for **controlled access** will be established.
- Embargoed data will include **visible metadata** and **conditions for future access**, and the reason and duration of the embargo will be documented.
- Publications will be made **available via open access channels**, and underlying datasets will be deposited in repositories at the time of publication, ensuring reproducibility.

3.4 DMP quality, review process & data inventory

An internal quality evaluation and reporting process is implemented throughout the entire project to assess both the project's data/products and its processes (see *Project Management Plan, Section 3.1 Internal Communication*). The detailed procedure for deliverable review is outlined in the *Project Management Plan*, *Section 5.2 Workflow for Deliverable Preparation*, including a peer-review process.

Throughout the project, data will be continuously analyzed and collected. To support this, the Project Management Team (PMT), comprising the Project Coordinator and Project Manager, will maintain and regularly update a **Scientific Publication Tracking Table** (see *Annex V*). This table will record proposals for manuscripts approved by the Executive Committee (ExCom) and will include the following mandatory and optional information:

Mandatory Information (for all papers, including drafts):

- Lead authors
- Co-authors and contributors (if already identified)
- Type of publication (e.g., empirical article, review)
- Target journal
- Related Work Package(s) and Task(s)
- Pre-print planned to be deposited on a repository platform (yes/no, and which platform)
- Estimated submission date

Optional Information before publication, but mandatory after acceptance:

- DOI
- Open science compliance status

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Name of the journal
- Acknowledgment compliance

This *Publication Tracking Table* will also be requested for completion during the Internal Progress Report, every 6 months (see *Project Management Plan, Section 3.1 Internal Communication*).

The **DMP** will be updated at **M34** and **M60** to reflect the progress and evolving needs of the project. This will include revisions to all sections of the DMP, as necessary, based on the ongoing developments in data collection, generation, and sharing within the online research data repository. Any new beneficiary, affiliated entity or partner joining the project is required to comply with the DMP. It will be updated in a timely manner to reflect such changes and ensure that all data management obligations are clearly communicated and followed.

Data Documentation for Re-use and Validation

Documentation to support data validation and re-use will be progressively developed throughout the project. This may include readme files, brief methodological notes, and other supporting materials, depending on the dataset. When relevant, metadata standards and controlled vocabularies such as those referenced in the INRAE Thesaurus will also support interoperability. Documentation will be stored alongside datasets in repositories like Zenodo, ensuring that essential context and licensing information (including license and author files) is available for future use.

Data Quality Assurance Processes

Data quality assurance processes will be defined progressively during the project, as they largely depend on the type of data generated and the methods used for their acquisition. Each WP will be responsible for proposing appropriate quality control measures. These may include the use of version control, clear documentation practices, validation procedures, and human review of data entries (e.g. checking questionnaire responses for anomalies or completeness). Depending on the nature of the data, additional checks may be carried out—such as ensuring that standardised acquisition methods are applied. For some data types, especially those collected through calibrated instruments, defining such processes is more straightforward. In all cases, common principles such as harmonisation of formats, metadata quality checks, and peer verification steps will help ensure coherence and reliability across the consortium.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

4. Legal and ethical requirements, codes of conduct

4.1 Legal issues, including intellectual property rights (IPR) and ownership

All aspects related to **General Data Protection Regulation** (GDPR), intellectual property, data ownership, and reuse rights are detailed in the **Grant Agreement** and the **Consortium Agreement**.

4.2 Ethical issues and codes of conduct

Ethical considerations are addressed in the D11.2 **Ethics Management Plan**, which outlines procedures to ensure compliance with applicable EU and national ethical standards. An **Ethics Advisory Board** (specifically for the YEAH! project) is currently being established to monitor and advise on ethical matters throughout the project lifecycle, particularly in cases where sensitive or unforeseen ethical issues may arise.

4.3 General Data Protection Regulation (GDPR)

The YEAH! project is fully compliant with the **General Data Protection Regulation** (GDPR) as stipulated in 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 and respects regulations on intellectual property rights (IPR).

If a new user wishes to access the YEAH! project's ShareDocs platform (secure internal document-sharing platform used exclusively by project members), they must register for an account. During registration, the user will be required to provide their email address and give consent for it to be processed and used for account-related communications. No other personal data is requested for account creation. Access to ShareDocs is restricted to project partners and collaborators; the platform is not accessible to external researchers or the general public. Users may withdraw their consent at any time, resulting in the immediate revocation of their access to ShareDocs.

If personal data are processed, compliance with the **General Data Protection Regulation** (GDPR, Regulation (EU) 2016/679) will be strictly ensured. Each partner will be responsible for applying GDPR within their national legal framework. Personal data will be pseudonymised or anonymised when possible, and stored securely on institutional servers or in the central repository with controlled access. By default, the **Data Protection Officer** of each institution will be consulted prior to any processing of personal data, so they can assess whether their

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

involvement is required and ensure compliance. Access to sensitive data will be restricted and traceable.

In addition, the following responsibilities and principles will be respected, particularly regarding the processing of health data and other sensitive personal data:

- Ensure that the purpose of processing is lawful and complies with the principles outlined in the GDPR: a clearly defined and legitimate purpose, the existence of a legal basis (such as the explicit consent of the data subject or another lawful basis recognised for processing sensitive data), minimisation of data to what is strictly necessary for the research, and respect for the principle of transparency toward data subjects.
- Guarantee the security of personal data by implementing appropriate technical and organisational measures (e.g., anonymisation or pseudonymisation) to prevent unauthorised access, loss, or accidental disclosure.
- Ensure compliance with the principles of *Privacy by Design* and *Privacy by Default* when developing new tools or processes involving health data.
- Clearly inform data subjects about the collection, use, storage, and sharing of their personal and health data, as well as their rights (access, rectification, deletion, portability, restriction of processing, etc.).
- Obtain informed consent when consent is the legal basis for processing personal and health data. This consent must be freely given, specific, informed, and unambiguous. Participants must also be clearly informed of their right to withdraw consent at any time, and this process must be straightforward.
- Ensure traceability of data by maintaining records of processing activities and enabling audits and reviews of data processing procedures to guarantee ongoing legal and ethical compliance.
- Implement specific measures for the handling of sensitive data such as health data, in accordance with the GDPR's enhanced protection requirements for this category of data.
- Comply with data retention obligations by ensuring personal and health data is kept only for as long as necessary for the research, and that it is securely deleted once that period has ended.
- Respect national legal frameworks regarding the processing of personal and health data, including country-specific data protection laws and health data-specific regulations.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

4.4 Sensitive Data

According to **Article 9 of the General Data Protection Regulation** (GDPR, Regulation (EU) 2016/679), sensitive data are a specific category of personal data that require heightened protection. These include:

- Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- Trade union membership;
- Genetic data and biometric data used to uniquely identify a natural person;
- Data concerning health;
- Data concerning a person's sex life or sexual orientation.

As a general rule, the collection and use of such sensitive data are prohibited under the GDPR, except in specific cases, such as:

- When the data subject has given explicit consent (through a free, specific, informed, and unambiguous act, preferably in writing);
- When the data has been clearly made public by the data subject;
- When processing is necessary to protect the vital interests of the data subject or another person;
- When processing is necessary for reasons of public interest and authorised by the relevant Data Protection Authority (e.g., CNIL);
- When the data concerns members or participants of non-profit organisations with political, religious, philosophical, or trade union aims, under strict conditions.

The proper management of sensitive data is imperative to maintain the individual privacy and remain in compliance with both EU and international regulations.

In order to ensure sensitive data is properly managed, data that is considered sensitive should first be identified. By principle, the processing of sensitive data, such as health data, is prohibited under Article 9 of the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679), unless it is justified by one of the exceptions listed in that article.

Each processing operation involving personal data must be based on a valid legal basis. For sensitive data, the justification must fall within the exceptions listed in Article 9 of the GDPR. If no valid exception applies and if the consent obtained does not meet the strict conditions set out in **Article 7** of the GDPR, then the processing is deemed unlawful.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

It is important to highlight that consent must be explicit, informed, freely given, specific, and unambiguous. Additionally, consent to the processing of personal data must be distinct from consent to participate in a research study. Furthermore, consent is not always the most appropriate legal basis, particularly when data are reused for secondary purposes (e.g., in the context of indirect data collection from data warehouses).

In order to determine the correct legal basis and applicable regulatory framework (e.g., conditions to be fulfilled, prior authorisations required), it is essential to precisely identify the types of personal data being processed and the purposes of the processing.

Sensitive data will be stored securely and only by the partners directly handling such data. Data protection measures will be implemented to safeguard the privacy of individuals. When a partner expects to process sensitive data, they must notify the Consortium and consult their institutional DPO, who is responsible for verifying whether their involvement is required.

Finally, national regulations may impose additional requirements on the processing of health-related data for research purposes. These conditions and procedures may vary across Member States, and it is necessary to take them into account when planning the data processing activities.

Relevant GDPR articles include:

- **Article 4**: Definitions of key GDPR terms;
- **Article 6**: Lawfulness of processing;
- **Article 7**: Conditions for valid consent;
- **Article 9**: Special categories of personal data and conditions for their processing.

Sensitive data will be stored privately by partners handling the data using robust technical and organizational measures to ensure security and confidentiality. These measures include, for example, encryption solutions such as Cryptomator, the use of password-protected computers and hard drives, and access controls to limit data accessibility. Sensitive data will not be published; instead, access to these datasets may be granted to authorized reviewers upon request and under strict conditions defined in the data processing agreements and in compliance with applicable data protection laws. These agreements will specify the purpose of data access, ensuring it aligns with the consent provided by participants and with the commitments made under GDPR and applicable national legislation. In any instance where sensitive data are collected, partners must notify

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

the Consortium, provide appropriate documentation demonstrating compliance with GDPR and applicable national regulations, and formally accept responsibility for the handling of such data.

4.5 Clinical Data

The clinical data generated within the YEAH! project will be handled with strict adherence to ethical guidelines and regulatory frameworks, particularly the **GDPR** and **ethical standards** for clinical research. It is essential to ensure that the processing of health data within the context of clinical research strictly complies with the principles of confidentiality, security, and regulatory compliance, including the GDPR. Each participating country may have specific legal requirements regarding the processing of sensitive data, and each partner will be responsible for ensuring that their practices comply with the legislative and ethical frameworks of their own country. All actors involved in processing these data must commit to respecting these legal and ethical obligations and assume full responsibility in case of non-compliance. All data collected during clinical trials will be pseudonymized to ensure participant confidentiality and privacy. Informed consent will be obtained from all participants, ensuring they are fully aware of how their data will be used, stored, and shared. Data will be securely stored using encrypted systems, and access will be restricted to authorized personnel only. Ethical oversight will be maintained through collaboration with ethics committees, ensuring compliance with both national and European standards for human research, including provisions for sensitive data and vulnerable populations.

The YEAH! consortium will also ensure full transparency and public access to all clinical results generated during the project (ClinicalTrials.gov), ensuring that these results are collected and reported to, on the one hand, maximise their usefulness to the scientific/technological community, and, on the other hand, allow for stakeholders (patients, healthcare professionals, public entities, etc) to find information about the therapeutic and diagnosis solutions they might be utilising. As part of this project, a processing of personal data will be implemented in a strict confidential manner to allow analysis of the results of the research. A data anonymization process will be applied to avoid revealing personal identities, and it is essential to specify that no data or results will be published before the data has been anonymized. It is therefore crucial that each actor formally commits to respecting these requirements, ensuring accountability in case of non-compliance. YEAH! will also ensure pre-registration of protocols and hypotheses for all meta analyses, epidemiological studies, and data collection studies (see *Grant Agreement, Clinical Studies section 5, §3.4*).

More details on the management of clinical data are reported in the Clinical Annex (see *GA*).

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

5. Other research outputs

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

Partners 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 development in **WP3** will be covered by a **Software Management Plan**, currently under preparation, which outlines the principles and processes for software management, sharing, and re-use. More details are provided in *Annex IV*.

6. Allocation of resources

6.1 Costs

Expenses related to providing open access to research data in Horizon Europe are **eligible for reimbursement**, in accordance with the conditions outlined in the A*rticle 17 of the Grant Agreement* and the relevant provisions for the specific cost category. Each project partner is responsible for planning the costs incurred in making data accessible to third parties beyond the consortium.

Costs associated with ensuring that data are FAIR (Findable, Accessible, Interoperable, Reusable) may include:

- Charges for **publishing scientific articles** that contain research data in Gold OA journals. Only publication fees in full open access venues for scientific publications are eligible for reimbursement. In the case of co-authored publications, the sharing of publication fees should be agreed upon by the authors on a case-by-case basis;
- Expenses related to the creation and maintenance of the **project website** are included in the budget allocated to the project coordinator (UGA).
- **Storage and archiving** of data on platforms such as Zenodo or other online repositories, which are currently free of charge; However, the cost of data storage may **vary depending on the volume of data** generated throughout the project;

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Application of Creative Commons licenses for copyright, which are also free of charge.

Each beneficiary is responsible for managing the data they generate. Any publication-related OA costs will fall under the responsibility of the partner(s) owning the data.

At this stage, no costs are foreseen for the long-term storage of open data in Zenodo. Further cost details and clarifications will be provided in subsequent updates of the Data Management Plan, if necessary.

6.2 Responsibility

Data management is a key activity that involves the entire project and must be coordinated and monitored both at the project and WP levels. It is also closely linked to the dissemination of project results, including publications.

The overall responsibility for data management rests with the **Project Coordinator** (UGA). However, each **Work Package Leader** (WPL) is accountable for managing the dataset and research data generated within their respective WP. The Project Coordinator (data manager) retains oversight of the entire project's data management, while the WPLs (WP data managers) ensure proper management of the data specific to their WPs.

Responsibilities related to the **processing of personal data** (GDPR) are specifically outlined in the **GDPR Appendix** (attached to the Consortium Agreement and available in *Annex VII*) and **completed by each WPL** at the start of their activities.

The Project Coordinator (data manager) and the WPLs (WP data managers) have the following responsibilities:

- Collaborating with the Project Management Team (PMT) to develop and implement the Data Management Plan (DMP) and related policies.
- Monitoring all data management activities, including data collection and publication, and ensuring adherence to deadlines.
- Ensuring that open research results (data and software) are deposited in the Zenodo community or an alternative OpenAIRE-compliant repository.
- Providing guidance and sending reminders to all project partners regarding data management tasks.
- Writing, updating, and uploading the DMP to the appropriate platform.
- Offering assistance in selecting the appropriate publication route (green or gold open access).
- Providing tailored support and further guidance on publishing scientific articles.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Ensuring that the OA policy of the chosen journal aligns with Horizon Europe open data requirements before submission.
- Monitoring that self-archived (green access) publications are uploaded to repositories and sending reminders to partners to ensure compliance.
- Ensuring that publication metadata is made available through the R&I Participant Portal (preferably via Zenodo) and by including a dedicated link on the YEAH! project website that redirects to the Zenodo community page where all open access results are deposited.
- Ensuring that research data linked to publications are available in repositories and properly connected to the corresponding publications.
- Monitoring embargo periods and issuing reminders to partners when necessary.
- Describing research data and publications according to the YEAH! Data Management policy, using appropriate metadata standards provided by the project.

6.3 Long term preservation

Given the volume of data collected and produced, a specific **ad-hoc secured central repository** for pseudonymised data and hosted in France (when in accordance with national laws and regulations) will be maintained by the **Project Coordinator** (UGA). The platform used should be SUMMER, a secure data environment designed to facilitate internal project collaboration and data management, but not intended for public data dissemination. This repository will remain operative **5 years after the project ends**, except for cohort data (WP1) which will stay with their owners (in this case data will be shared only at the level of pooled results).

7. Data security and storage

The following guidelines will be followed in order to ensure the security of the data:

- Store data in at least two separate locations to avoid loss of data. The SUMMER infrastructure includes three data centers, ensuring redundant backups across two geographically distant sites.
- Encrypt data if it is deemed necessary by the participating researchers;
- Limit the use of USB flash drives and other mobile storage devices whenever possible. When such devices are used (e.g., USB keys, external hard drives), they must be fully encrypted to prevent unauthorized access in case of loss or theft.

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Label files in a systematically structured way in order to ensure the coherence of the final dataset.

As an initial step, only the Consortium Partners will have access to the repository where dataset and metadata are filed. The protection of data will be ensured through procedures and appropriate technologies (e.g., HTTPS protocol for the encryption of all internet transactions and appropriate European and Internet security standards from ISO, ITU, W3C, IETF and ETSI). If data will be kept in a certified repository, then the security standards of that repository will apply.

For data storage, we are considering the use of **SUMMER**, though this has not yet been confirmed. Final decisions on the storage solution will be made in due course, taking into account the project's evolving needs and requirements.

Following, scientific publications and articles, the dataset deliverables and the final demonstrator research results will be shared through **Zenodo** and other database to promote the data making FAIR.

Data preservation

Data preservation refers to data upkeep and maintenance to ensure that the integrity of the data is upheld in the future. This includes properly maintaining the data repository and data backups to ensure the long-term value of the data.

Data backups are expected to occur once a week, though this is subject to change based on the amount of data that will ultimately need to be backed up. In the case of data stored on the SUMMER infrastructure, backups are performed nightly, with the possibility of retrieving previous versions of data from the past 30 days, providing an additional layer of security for long-term access and recovery.

Estimating the **size of the data** and its access frequency is difficult to accomplish at this time since the data that will be initially included in the project has not been fully identified yet. An initial inventory will be identified by the Consortium. A better impression on the size of the data can be made after these datasets have been identified. However, YEAH! is designed and intended to have useful data added on a consistent basis, so the size will likely change over time, as well.

The data can be downloaded and copied to **personal work computers** by each partner having access to the project repository, without automatic monitoring. In this case, partners must ensure that appropriate data protection measures are in place on their personal work computers. This includes:

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Using encrypted devices and/or folders for storing sensitive data
- Protecting access to the computer with strong passwords and/or two-factor authentication
- Ensuring up-to-date antivirus and firewall software
- Not transferring data to unauthorized devices or external storage without prior security checks
- Deleting local copies of data once they are no longer necessary or after the end of the project

Each partner is responsible for ensuring compliance with data protection rules as per the GDPR and relevant national legislation.

For the clinical data, on REDCap each partner has access to its own data but not to the data of the all the partners unless otherwise specified

Data Security and Long-term Storage Provisions

At this stage, we acknowledge the importance of defining clear provisions for data security, long-term storage, and backup. However, the specifics have **not yet been fully determined**.

For data security, we are aware of the need to implement appropriate measures to safeguard data throughout the project's lifecycle. This includes ensuring that data is safely stored in trusted repositories for long-term preservation and curation.

Additionally, we recognize the importance of outlining how data will be stored and backed up during the research process, ensuring its integrity and availability. These provisions will be **further developed and refined as the project progresses** and as the storage solutions are finalized.

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

8. Annex I – Digital Curation Centre (DCC) – Information

The Digital Curation Centre (DCC) was launched in March 2004 with support from Jisc and the Engineering and Physical Sciences Research Council (EPSRC) to help solve digital curation and longer-term preservation challenges that could not be tackled effectively by any single institution or discipline. They produced guidelines to help organisations make the best use of data in and for research. The content presented in this section can serve as a reference framework for the effective implementation of the Data Management Plan (DMP).

8.1 Data Description

- Give a summary of the data you will collect or create, noting the content, coverage and data type, e.g., tabular data, survey data, experimental measurements, models, software, audiovisual data, physical samples, etc.
- Consider how your data could complement and integrate with existing data, or whether there are any existing data or methods that you could reuse.
- Indicate which data are of long-term value and should be shared and/or preserved.
- If purchasing or reusing existing data, explain how issues such as copyright and IPR have been addressed. You should aim to minimise any restrictions on the reuse (and subsequent sharing) of third-party data.

8.2 Data Format

- Clearly note what format(s) your data will be in, e.g., plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw).
- Explain why you have chosen certain formats. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community.
- Using standardised, interchangeable or open formats ensures the long-term usability of data; these are recommended for sharing and archiving.

8.3 Data Volume

- Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports).
- Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs?
- Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

8.4 Data Collection

- Outline how the data will be collected and processed. This should cover relevant standards or methods, quality assurance and data organisation.
- Indicate how the data will be organised during the project, mentioning, e.g., naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse.
- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies.

8.5 Metadata and Documentation

- What metadata will be provided to help others identify and discover the data?
- Researchers are strongly encouraged to use community metadata standards where these are in place. Data repositories may also provide guidance about appropriate metadata standards.
- Consider what other documentation is needed to enable reuse. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data and software used to collect and/or process the data.
- Consider how you will capture this information and where it will be recorded, e.g., in a database with links to each item, in a 'readme' text file, in file headers, etc.

8.6 Ethics and Privacy

- Investigators carrying out research involving human participants should request consent to preserve and share the data. Do not just ask for permission to use the data in your study or make unnecessary promises to delete it at the end.
- Consider how you will protect the identity of participants, e.g., via anonymisation or using managed access procedures.
- Ethical issues may affect how you store and transfer data, who can see/use it and how long it is kept. You should demonstrate that you are aware of this and have planned accordingly.

8.7Storage and Security

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- Describe where the data will be stored and backed up during the course of research activities. This may vary if you are doing fieldwork or working across multiple sites so explain each procedure.
- Identify who will be responsible for backup and how often this will be performed. The use of robust, managed storage with automatic backup, for example, that provided by university IT teams, is preferable. Storing data on laptops, computer hard drives or external storage devices alone is very risky.
- Also consider data security, particularly if your data is sensitive e.g., detailed personal data, politically sensitive information or trade secrets. Note the main risks and how these will be managed. Also note whether any institutional data security policies are in place.
- Identify any formal standards that you will comply with, e.g., ISO 27001.

8.8 Roles and Responsibilities

- Outline the roles and responsibilities for all activities, e.g., data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Individuals should be named where possible.
- For collaborative projects you should explain the coordination of data management responsibilities across partners.

8.9 Preservation

- Outline the plans for data sharing and preservation - how long will the data be retained and where will it be archived? Will additional resources be needed to prepare data for deposit or meet any charges from data repositories?

9. Annex II: Datasets collected in the YEAH! Project

Dataset reference and name: XX – WPX Organisation in charge: XXX		
Data Type	**Data Standards – Formats**	**Data Generation**
Estimated Data Size	**Data Sharing**	**Storage and Preservation**

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

10. Annex III: Datasets that could be made openly accessible

Data Producer	Brief Description of the Dataset	Foresee use and re-use	Possibility to share the dataset beyond the consortium	Publication associated

11. Annex IV: Software Management Plan

The development work in WP3 will be addressed through a dedicated Software Management Plan (SMP), which is currently under preparation. This SMP will be delivered in a future version of the DMP. It will outline the principles and processes for software development, documentation, licensing, sharing, and long-term preservation. The plan will be aligned with the FAIR principles, ensuring software findability, accessibility, interoperability, and reusability. To ensure consistency and quality, the SMP will build upon the methodology and best practices outlined in the PreSoft project.

12. Annex V: Scientific Publication Tracking Table

Reporting	Related WP(s) and Task(s)	Title	Date Published or Estimated Date of Publication	(Target) Journal	Type of publication (empirical article, review, etc.)
Lead Author(s) + co-authors	**DOI (if available)/Link**	**Open Acces Y/N**	**Status (draft/in review/published)**	**Acknowledgment Compliance (EU funding statement + disclaimer + institutional statements)**	**Pre-print planned to be deposited on a repository platform (Y/N and if Y, which platform?)**

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

13. Annex VI: COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY (— ARTICLE 17) (HE)

From *Annotated Grant Agreement: V1.0– 01.05.2024*

13.1 Dissemination

Dissemination of results

The beneficiaries must disseminate their results as soon as feasible, in a publicly available format, subject to any restrictions due to the protection of intellectual property, security rules or legitimate interests.

A beneficiary that intends to disseminate its results must give at least 15 days advance notice to the other beneficiaries (unless agreed otherwise), together with sufficient information on the results it will disseminate.

Any other beneficiary may object within (unless agreed otherwise) 15 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the results may not be disseminated unless appropriate steps are taken to safeguard those interests.

Additional dissemination obligations

Where the call conditions impose additional dissemination obligations, the beneficiaries must also comply with those.

13.2 Open Science

Open science: open access to scientific publications

The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications
- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or a license with equivalent rights; for monographs and other

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

long-text formats, the license may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and
- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

Only publication fees in full open access venues for scientific publications are eligible for reimbursement.

Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- establish a data management plan ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements
- as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a license/dedication with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
- be against the beneficiary's legitimate interests, including regarding commercial exploitation, or

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

- be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP
- provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.

Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s) and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs.

Open science: additional practices

Where the call conditions impose additional obligations regarding open science practices, the beneficiaries must also comply with those.

Where the call conditions impose additional obligations regarding the validation of scientific publications,the beneficiaries must provide (digital or physical) access to data or other results needed for validation of the conclusions of scientific publications, to the extent that their legitimate interests or constraints are safeguarded (and unless they already provided (open) access at publication).

Where the call conditions impose additional open science obligations in case of a public emergency, the beneficiaries must (ifrequested by the granting authority) immediately deposit any research output in a trusted repository and provide open access to it under a CC BY licence, a Public Domain Dedication (CC 0) or equivalent. As an exception, if the access would be against the beneficiaries' legitimate interests, the beneficiaries must grant non-exclusive licenses —under fair and reasonable conditions —to legal entities that need the research output to address the public emergency and commit to rapidly and broadly exploit the resulting products and services at fair and reasonable conditions. This provision applies up to four years after the end of the action (see Data Sheet, Point 1).

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

13.3 Plan for the exploitation and dissemination of results including communication activities

Unless excluded by the call conditions, the beneficiaries must provide and regularly update a plan for the exploitation and dissemination of results including communication activities.

14. Annex VII: GDPR Appendix

This annex must be completed and sent to the coordinator before the start of each WP by the parties participating in the WP, and if personal data is processed or exchanged within the framework of the WP.

DESCRIPTION OF THE PROCESSING OPERATIONS FOR WHICH THE PARTIES ARE JOINT CONTROLLERS AND DIVISION OF TASKS

Legal status of Research:

- Treatment description:
- Partners carry out the following Processing operations (please indicate the name of the Data Controller):
- The purpose(s) of the processing:
- Legal basis(s) for processing:
- Categories of Data processed:
- Contact details for each Partner's DPO:
- List of subcontrators, if any:
- Existence of data transfers outside the EU/EEA:
- Shelf life of the Data:

Technical specifications:

- Solutions used for data storage:
- Method and location of data storage:
- Transfer method used between the Parties:

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Tasks / roles	Parties	Comments
If there is a legal source, who designates it as the data controller? - Which party is the promoter?		
Which Party initiates the research?		
Which Party constructed the research protocol (i.e. determined the means and ends)? - Adapting the protocol specifically for national regulatory submissions is not considered "construction". - Minor contributions to the protocol are not considered "construction".		
If the project involves the creation of a database for one or more secondary uses, which Party has effective control over this database?		
Which Party is entitled to engage third parties (including processors) to participate in the processing?		

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Tasks / roles	Parties	Comments
Which Party is providing the financial resources for the Treatment? - Is a Party financing the project?		
Which Party is responsible for drafting the privacy impact assessment (PIA) - Article 35 of the GDPR?		
Which Party is responsible for regulatory procedures concerning the processing of personal data?		
Which Party is responsible for obtaining Data from the persons concerned (participants)?		
Which Party is responsible for the analysis (or other operations of an intellectual nature) of the Data?		
Which Party is responsible for notifying the supervisory authority in the event of a data breach (and who decides when and how to notify)?		
Which Party is responsible for notifying Data Subjects in the		

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Tasks / roles	Parties	Comments
event of a data breach (and decides when and how to notify)?		
Which Party is responsible for data security, integrity and confidentiality? - A data provider is not subject to security, integrity and confidentiality obligations for the data it has collected, with the exception of providing a "state of the art" transfer method.		
Which Party is responsible for ensuring that the treatment is carried out in accordance with the research protocol and established scientific practices?		
Which Party is responsible for ensuring that processors provide the required level of data protection (in accordance with Article 28 of the GDPR).		
Which Party must be displayed / announced as an access point for data subjects to exercise their rights (RGPD)?		

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977

Tasks / roles	Parties	Comments
Which Party is responsible for providing the information media (in accordance with Articles 13 and/or 14 of the RGPD)?		
Which party is responsible for writing the information material?		
Which Party is responsible for managing requests from Data Subjects to exercise their rights?		
Which Party provides the Data transfer method?		

Funded by the European Union

This project has received funding from the European Union´s Horizon Europe under grant agreement No 101155977