Plan Overview

A Data Management Plan created using DMPonline

Title: Evaluating the feasibility and acceptability of the Healthbox for metabolic syndrome: a pilot study

Creator: Aura Vernooy-Geerling

Principal Investigator: Niels H. Chavannes

Affiliation: Leiden University Medical Center

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: LUMC data management plan

ORCID iD: 0000-0002-8607-9199

Project abstract:

Metabolic syndrome (MetS) substantially increases the risk of chronic conditions such as cardiovascular disease and type 2 diabetes. Approximately 30% of the Dutch population meets the criteria for MetS, with a disproportionately high prevalence among individuals with lower socioeconomic status (SES). The HealthBox is a personalized, home based eHealth lifestyle intervention designed to support sustainable behavior change. The intervention integrates three key components: (1) the Ancora app — a digital platform that provides participants with access to a lifestyle program focused on nutrition, physical activity, and stress management; (2) the Box — a physical kit containing a blood pressure monitor, weighing scale, and activity tracker, all connected to the Ancora platform to enable continuous telemonitoring; and (3) the Recommender System — A digital classification system that assigns participants to a specific profile and tailored intervention pathway within the Ancora program based on their characteristics and preferences. This ensures that each participant receives a personalized lifestyle program. Before initiating a large-scale implementation through a stepped-wedge hybrid type 2 trial, it is essential to evaluate the feasibility and acceptability of the intervention in a pilot study.

ID: 181010

Start date: 01-01-2025

End date: 01-05-2026

Last modified: 23-07-2025

imply that the creator(s) endorse, or have any relationship to, your project or proposal

Evaluating the feasibility and acceptability of the Healthbox for metabolic syndrome: a pilot study

Be aware!

This LUMC DMP is approved by ZonMw, NWO, and can be used for Horizon Europe and ERC projects. It is also approved for the Health-Holland projects that are monitored by the LUMC.

!! For the Health-Holland projects that are coordinated by Health-Holland, you need to use the template provided by Health-Holland.

If you have doubts on which template to use, please contact one of the data stewards via rsd@lumc.nl

n.a.

I. General information

I.1 Name of researcher responsible for DMP

Aura Vernooy-Geerling

I.2 Department of researcher responsible for DMP

Public Health and Eerstelijnsgeneeskunde (PHEG)

I.3 ORCID ID of researcher(s) responsible for DMP

0009-0006-5682-7113

I.4 Supervisor(s) of project, if applicable

Prof. Dr. N.H. Chavannes
Dr. M.J.J. van der Kleij-van de Sluis
Dr. M.J. Kasteleyn

I.5 If applicable for your study or project, please provide:

If one or more numbers are not applicable for your study or project, please add '/ ' in the appropriate text box.

ABR number	'/ '
METC number	'/ '
EudraCT number	'/ '
GMO number	'/ '
CCD application number	'/ '

I.6 If applicable: list the partner organisation(s) involved in your study or project.

University of Twente
Delft University of Technology
University of applied sciences Utrecht
National eHealth Living Lab (NeLL)
Ancora Health B.V.

Partners involved without involvement with data collection and data management:

Diabetes Fonds
Patiëntenfederatie Nederland
Stichting Pharos

II. About this DMP

II.1 Date of first DMP version

2025-06-26

II.2 Consulted LUMC data stewardship expert(s)

Note: This field is a requirement for most funders. DMPs will only be reviewed by funders after you have consulted one of the LUMC data stewardship experts.

To request for review: see section 6.

The LUMC has both local DMP supporters at the departments and central data stewardship experts, both specialized in clinical data stewardship and translational and fundamental research. This DMP was reviewed by:

Name	
Telephone number	
E-mail address	
Date of consultation (dd/mm/yyyy)	

Name	
Telephone number	
E-mail address	
Date of consultation (dd/mm/yyyy)	

II.3 Changes made to an earlier version of this DMP

Part of DMP	Date of change (in dd/mm/yyyy)	Question number(s)	Adaptation(s) made
I. General information			
II. About this DMP			
1. Data collection			
2. Data documentation			
3. Data storage during your study/project			
4. Archiving data within the LUMC			
5. Data publication & reuse			

Remarks: n.v.t.

1. Data collection

1.1 What type of study or project will you conduct?

- Other study/project

Study/project with human participants: It is a non-WMO pilot study with 30 participants that evaluate the feasibility and acceptability of a eHealth lifestyle intervention. No data from HiX will be used.

1.2 Is your informed consent form according to the *LUMC-based CCMO model form*?

- Yes

1.3 Describe what is mentioned in the informed consent about data sharing and re-use.

Participants are informed that their data will be handled according to the FAIR principles, meaning that anonymised research data may be made Findable, Accessible, Interoperable and Reusable for future scientific purposes, in line with open science practices.

However, we also explain that participants' privacy and confidentiality are of utmost importance. Therefore, any data that could directly or indirectly identify them will be **pseudonymised or anonymised** before being shared or reused. Personally identifiable information (PII), such as names, contact details, or unique codes, will never be made publicly available or shared outside the research team without explicit additional consent.

Participants are told that their anonymised data may be reused by other researchers for related studies, meta-analyses, or secondary research, but only under conditions that ensure secure and ethical use. These may include data access committees, data use agreements, or repositories with controlled access.

We make clear that participation is voluntary and that individuals can withdraw their consent at any point without consequences for their care or relationship with the researchers. If they choose to withdraw, their data will no longer be used unless it has already been anonymised and incorporated into analyses in a way that cannot be traced back to them.

1.4 How do you ensure that participants, who have withdrawn their informed consent, are removed from the data and thus are not available for reuse? Do you have a procedure in place for this?

When a patient withdraws from the study, the corresponding data will be removed from Castor EDC and any files that are stored in a DataSafe.

1.5 Describe what you will do to pseudonymize or anonymize your data: How will you pseudonymize, where will identifiable data be stored and who is responsible for managing this data during the study or project?

All questionnaires and informed consent forms (ICFs) for stakeholders will preferably be administered digitally. The screening questionnaire, ICFs and health-and-behavior questionnaire will be completed prior to the start of the pilot HealthBox intervention. All ICFs will be securely stored at the Department of Public Health and Primary Care of the LUMC.

Audio recordings of interviews, as well as the digitalized sociodemographic data, will be processed confidentially and securely stored on a protected server at the same department until transcription is complete. Data will be stored in a digital database only accessible to the researchers.

To ensure participant privacy, both the qualitative transcripts and sociodemographic data will be pseudonymized using unique codes. A separate key will be used to link these codes to individual participants, and this key will be stored securely in a protected folder on the LUMC I:-drive.

All personal data will be handled in accordance with the General Data Protection Regulation (GDPR) and the Dutch Implementation Act (Uitvoeringswet AVG, UAVG). Access to the data will be restricted to authorized study personnel. Transcripts of the audio recordings will be retained for a period of 15 years.

1.6 Which data assets do you create during your study/project?

Main study parameters:

Feasibility of the Healthbox will be assessed by examining two key components:

- Technical feasibility; technology use and system usability.
- Clinical feasibility; participant experiences of integrating the intervention components, namely the Ancora app and home-monitoring devices, into their daily routine

Quantitative measurements:

- Feasibility of Intervention Measure (FIM)
  The FIM (Dutch version: HIM – Haalbaarheid van Interventie Meting) will be used to assess both technical and clinical feasibility—particularly how easily participants can incorporate the intervention into their everyday lives and health routines. The FIM is a 4-item instrument developed to assess perceived feasibility of a health intervention from the user's point of view. The items cover whether the intervention seems implementable, doable, practical, and easy to use in a real-world context. Responses are recorded on a 5-point Likert scale ranging from “completely disagree” to “completely agree.” The total score provides a quantitative reflection of perceived feasibility, with higher scores indicating greater feasibility. The Dutch version (HIM) has been validated for use in Dutch-speaking populations.
- System Usability Scale (SUS)
  The System Usability Scale (SUS) is a widely used, reliable 10-item questionnaire that assesses the usability of a system from the user's perspective. In this study, the SUS will be used to evaluate technical feasibility, in specific the participants’ subjective assessment of the usability of the eHealth platform (Ancora app) and the associated home-monitoring tools technical feasibility. In the SUS, each item is rated on a 5-point Likert scale ranging from “Strongly disagree” to “Strongly agree.” The items alternate between positive and negative formulations to reduce response bias. The SUS provides a single score ranging from 0 to 100, with higher scores indicating better usability. A score above 68 is generally considered above average usability. In this study, the SUS will be used to evaluate participants’ subjective assessment of the usability of the eHealth platform (Ancora app) and the associated home-monitoring tools.

Qualitative measurements:

- Think aloud sessions
  The think aloud sessions will be used to evaluate technical feasibility and will be conducted after the distribution and explanation of home-monitoring devices right before the start of the 6-week intervention period. The think-aloud script will prompt participants to verbalize their thoughts, expectations, and any difficulties they encounter while using the devices for the first time (Wolcott et al. 2021). The script will include standardized prompts such as "What are you thinking now?", "What do you expect to happen next?", and "Is anything unclear to you at this moment?" This method allows researchers to identify potential usability barriers, misunderstandings, or areas where further instructional support is needed. The insights gathered from these sessions will inform iterative refinements to the device instructions and support materials, ensuring that participants can confidently and effectively incorporate the home-monitoring devices into their daily routines.
- Focus groups
  In addition, semi-structured focus group discussions will be held after participants have engaged with the intervention for a 6-weeks intervention period to evaluate the technical and clinical feasibility of the intervention. These discussions aim to gather collective feedback on the intervention’s practicality and integration into daily routines. The focus groups will explore experiences with the content, design, and support features of the intervention, as well as any contextual factors influencing its use. Discussions will be recorded, transcribed verbatim, and analyzed using thematic content analysis.

Acceptability of the Healthbox will be assessed by examining two key components:

- User Satisfaction
- Expectations & Needs of the user

Quantitative measurements:

- Acceptability of Intervention Measure (AIM)
  The AIM is a validated, psychometrically robust instrument designed to assess the perceived

acceptability of a healthcare intervention. It consists of four items, each rated on a 5-point Likert scale ranging from “Completely disagree” to “Completely agree.” The items capture how agreeable, suitable, and satisfactory participants find the intervention overall. Higher scores indicate greater acceptability. Although the AIM does not explicitly distinguish between satisfaction and expectations/needs, its items collectively reflect both. Participants’ satisfaction is indicated by the degree to which they find the intervention appealing, whereas their expectations and needs are reflected in their approval of the intervention. A Dutch version of the AIM will be used to assess how participants subjectively evaluate the three pre-designed intervention trajectories recommended by the recommender system.

Qualitative measurements:

- Focus groups
  In addition to the AIM, qualitative data from the same focus group discussions used to evaluate feasibility will also be analyzed to further explore participants' expectations and the perceived alignment of the intervention with their personal needs. This mixed-methods approach enables a nuanced understanding of both the feasibility and acceptability of the intervention from the user's perspective.

Other study parameters:
Additional quantitative parameters include drop-out rate, log-in rate, retention rate, task success rates, error rates, and time on tasks. These usage-related data will be collected through backend monitoring of the Ancora app. These parameters provide objective insights into participant engagement and into the app usage. By triangulating these data points with qualitative insights from think-aloud sessions and focus groups, we aim to explore how such behavioral indicators relate to the feasibility and acceptability of the HealthBox intervention. This integrated approach helps identify specific barriers and facilitators that may influence participants’ sustained use and perceived value of the intervention.

Descriptive measures:
The following descriptive variables will be collected: health literacy, digital literacy, age, gender, income, and medication use. These variables are self-reported measurements to provide contextual information about the study population. Health and digital literacy are included because they may influence participants’ ability to navigate and benefit from the HealthBox intervention. Demographic characteristics such as age, gender, income, and medication use offer further insight into participant diversity and may help identify subgroups for whom the intervention is particularly suitable or needs adaptation.

1.7 Is one of the outcomes of your project software? You can think of scripts, modules, tools, an app, a analysis pipeline etc.

- No

1.8 How will you collect, create and/or capture your data? Briefly describe what you need to collect or access the data. Think about protocols, tools, equipment, hardware etc.

See answer 1.6

1.9 What is the size and format of your digital data? And what software do you need to

collect, process and analyse these data sets?

if you don't know the size estimate, you can give a range: < 1 GB, 1-10 GB, 10-100 GB, 100 GB - 1 TB, > 1 TB

Stage	Specification of data set	Software choice	File format	Data size estimate*
Data collection	1. Clinical measurement outcomes	Castor EDC	not applicable	not applicable
	2. Log files data	Ancora app	not applicable	not applicable
	3. Questionnaires	Castor EDC	not applicable	not applicable
	4. Think aloud outcomes	voice recorder	not applicable	not applicable
	5. Focus groups outcomes	voice recorder	not applicable	not applicable
Raw data	1. Clinical measurement outcomes	export to SPSS	.sav	< 1 Gb
	2. Log files data	export to SPSS	.sav	< 1 Gb
	3. Questionnaires	export to SPSS	.sav	< 1 Gb
	4. Think aloud outcomes	export to audio files and transcripts	.mp3 + .docx	0.2 GB audio + < 1 MB per transcript
	5. Focus groups outcomes	export to audio files and transcripts	.mp3 + .docx	0.2 GB audio + < 1 MB per transcript
Processed data	1. Clinical measurement outcomes	SPSS	.sav	< 500 MB
	2. Log files data	SPSS	.sav	< 500 MB
	3. Questionnaires	SPSS	.sav	< 500 MB
	4. Think aloud outcomes	Transcribed + Atlas.ti	.atlproj	< 500 MB
	5. Focus groups outcomes	Transcribed + Atlas.ti	.atlproj	< 500 MB
Results	1. Clinical measurement outcomes	SPSS output	.spv	< 50 MB
	2. Log files data	SPSS output	.spv	< 50 MB
	3. Questionnaires	SPSS output	.spv	< 50 MB
	4. Think aloud outcomes	Thematic summaries, tables, quotes	.xlsx	< 10 MB
	5. Focus groups outcomes	Thematic summaries, tables, quotes	.xlsx	< 10 MB
Other…				

1.10 What is the estimated *total size* of the digital data you generate in this study/project?

You can use the 'additional information' field for more details.

- 0-10 GB

Data collection:

- Clinical measurement outcomes: n.v.t.
- Questionnaires: n.v.t.
- Log files data: n.v.t.
- Think aloud (MP3): n.v.t.
- Focus groups (MP3): n.v.t.
- **Subtotal: 0 MB**

Raw data:

- Clinical measurement outcomes: < 1 GB
- Questionnaires: < 1 GB
- Log files data: < 1 GB
- Think aloud (MP3): ± 0.2 GB + ± 1 MB
- Focus groups (MP3): ± 0.2 GB + ± 2 MB
- **Subtotal: ± 3.4 GB**

Processed data:

- Clinical measurement outcomes: < 500 MB
- Questionnaires: < 500 MB
- Log files data: < 500 MB
- Think aloud (MP3): < 500 MB
- Focus groups (MP3): < 500 MB
- **Subtotal: < 2.5 GB**

Results:

- Clinical measurement outcomes: < 50 MB
- Questionnaires: < 50 MB
- Log files data: < 50 MB
- Think aloud (MP3): < 10 MB
- Focus groups (MP3): < 10 MB
- **Subtotal: < 170 MB**

Total:
± 6.0 GB

1.11 Are there any non-digital data or outputs that the project will generate?

- Yes

1.12 Please explain briefly what non-digital data or outputs you have:

1. Audio tapes from the think aloud and focus groups sessions.
2. Informed Consents (maybe, we prefer digital consents unless participants insist on filling out a

paper informed consent)

1.13 Will the project use existing data?

- No

2. Data documentation

2.1 How will files and folders be named and structured?

All files will be named with the naming convention: ‘yyyymmdd_[type]_[name]_[version]’

Date is date of creation. For each new version a new document is created.
Type can be: raw/clean/analysis/SOP/documentation
Name: initials of the creator
Raw data files will be locked by selecting the ‘read-only’ option in the file preferences. The suffix ‘_locked’ will be added to the file name and files will be stored in the ‘_raw’ folder. Obsolete files will be moved to an ‘Archive’ subfolder of the folder the document was in originally.
Proposed folder structure, example for I:-drive
[Project name] contains:

- [Projectname]_raw
- [Projectname]_cleaning
- [Projectname]_locked
- [Projectname]_analysis
- [Projectname]_SOP
- [Projectname]_documentation

2.2 How will versions and changes be handled?

They will be logged with a major change nr followed by a minor change nr. Example: v0.1. The 0 (major change nr.) means that the document is still in a "draft" stage. Ones the major change nr. is changed to 1, we will have a first definite version. The minor change nr. are their to keep up with the revisions made by collegues and myself.

2.3 Project level metadata: What general metadata (standard) will be used to describe the total of all data in your study/project ?
Please describe briefly how you will create this.

- Dublin Core

Not sure yet.

2.4 Data content level metadata: Which field-specific metadata standard(s) will be used to describe and/or standardize data and variables? Please describe briefly how you will create this.

We will not use any field-specific metadata standards to describe and standardize data and variables.

2.5 What supporting information and/or documentation will you create to enhance understanding of the data? Please describe briefly what is needed for peers to understand, work and/or reproduce the data.

The study protocol will be stored with the data after approval by the nWMO commission.

A data dictionary (code book) will be available for the questionnaires and clinical data. It will be added after export of the data from Castor for the questionnaires, and upon receipt of the clinical data. We'll have this for the Ancora log files too.

All syntaxes used in data cleaning and analysis (including annotation describing the goal of processing steps) will be stored to facilitate replication.

For qualitative data, coding frameworks and processing files from Atlas.ti (including codebooks, memos, and network structures) will be saved in exportable formats to ensure transparency and enable reuse.

A readme.txt with a list of all available files and a description of their contents will be created at the end of the project, before archiving the data. Lab journal entries and additional protocols will be exported as pdf. We will also include the necessary software and tools needed for reuse and state whether embargoes, licences, commercial objectives or other conditions (like stated in informed consent agreements) have been imposed on the reuse of data.

2.6 Please tick the box to confirm that you are aware of and adhere to the applicable rules and codes of conduct for your study or project:

- **General**
  - **VSNU Code of Conduct for Research Integrity**
  - **LUMC data management guidelines**
  - **LUMC privacy policy**
- **Human research:**
  - **General Data Protection Regulation (GDPR; in Dutch: AVG)**
  - **Medical Treatment Contracts Act (In Dutch: WGBO)**
  - **Medical Research Involving Human Subjects Act (In Dutch: WMO)**
  - **Quality Assurance for Research involving Human Subjects**
  - **Code of Conduct for Medical Research (e.g. GCP)**
  - **Code of Conduct Responsible Use of Human Tissue**
- **Non-human research:**
  - **Experiments on Animals Act**

Please add an explanation when needed in the 'additional information' field.

- I'm aware of and adhere to the rules and codes of conduct that are applicable for my study.

2.7 Indicate which permits you will acquire for your study and add an explanation when needed in the 'additional information' field:

- Report the collection of (in)directly identifiable (research) data to the Data Protection Officer

3. Data storage during your study/project

3.1 Where will you store the different parts of your digital data? Please describe the storage location for each dataset that you defined in the table in question 1.9. Examples of storage options are given in the guidance.

Data set/type	**Storage option**
Results and figures	I-drive or SharePoint Office 365
Documentation	I-drive or SharePoint Office 365
Key file(s)	DataSafe
Clinical data	Castor EDC exports in DataSafe or I-drive or SharePoint Office 365?
Questionnaires	Castor EDC exports in DataSafe or I-drive or SharePoint Office 365?
Log files data	exports in DataSafe or I-drive or SharePoint Office 365?
Think aloud data	audio tapes exports in DataSafe or I-drive or SharePoint Office 365?
Focus group data	audio tapes exports in DataSafe or I-drive or SharePoint Office 365?

3.2 Please describe how backup and availability are guaranteed for each part of your data during the study/project.

Backup policy for LUMC network drives:
Every day a copy is automatically made of any changed files on the I and J disks . A full backup of all data is made once per week. IT&DI also makes monthly and quarterly backups so that earlier data can also be restored if necessary. DataSafes are also included in the LUMC backup policy. We can restore files up to four weeks ago from I and J disks (including DataSafe) and for older versions we need to contact the IT helpdesk.

Microsoft Office 365 guarantees file availability and provides a roll-back to a version up to a month earlier.

Castor EDC backups are made four times a day and stored at another geographical location by Castor. The backup files are kept for fifteen days. Reserve copies can be restored but this might add additional costs.

3.3 For non-digital data you described in question 1.11 and 1.12, please specify briefly where you will store these non-digital data and describe who is responsible for handling and storage of these outputs.

audio tapes exports and scans of written informed consents will be stored in DataSafe or I-drive or SharePoint Office 365.

3.4 How will access to data be managed during the project?

Please specify for each storage device the tools and procedures that you use to ensure that only authorized persons have access to data. Outline roles and responsibilities for all activities during your project, e.g. data capture, metadata production, data quality, storage and backup. For collaborative projects you should explain the coordination of data management responsibilities across partners.

1. I-drive (LUMC secure network drive)

- Access control: Access to the I-drive is restricted to LUMC employees with appropriate rights. Folder-level permissions are managed by the departments application manager.
- Tools: Authentication is managed via LUMC login credentials and two-factor authentication. Use: The I-drive is primarily used for storing administrative documents, consent forms, and internal reports that are not shared outside the institution.
- Back-up: The I-drive is automatically backed up daily by LUMC’s ICT department.

2. SharePoint 365 Office (Microsoft Teams environment)

- Access control: SharePoint is used for collaborative work within the project team. Access is granted by the the departments application manager. External collaborators can only gain access via secure invitation and approval.
- Tools: Access is controlled via LUMC Microsoft accounts and requires multifactor authentication.
- Use: Used for real-time collaboration on documents, protocols, and planning. Data stored here is not sensitive or pseudonymized before upload.
- Back-up: Data stored in SharePoint is automatically backed up and version-controlled via Microsoft cloud infrastructure.

3. DataSafe (LUMC secure environment for research data)

- Access control: DataSafe is used to store sensitive, pseudonymized or coded research data. Access is limited to the research team, data analysts, and statisticians who are formally registered to the project. Access is role-based.
- Tools: Secure login, two-factor authentication, role-based access control, and audit trails to track data access and changes.
- Use: All identifiable or sensitive research data are stored and processed exclusively in DataSafe. This includes patient data collected via HealthBox and survey responses.

- Back-up: Automatic encrypted backups are performed daily by the LUMC IT department.

Roles during project:

- Principal Investigator (PI): Overall responsibility for data management, including access rights, compliance, and oversight.
- Data Steward: Ensures data is stored in the correct locations, access rights are up to date, and backups are performed.
- Researchers and PhD students: Responsible for data collection (e.g., survey or HealthBox data), ensuring metadata is added where relevant, and following Good Clinical Practice (GCP) and FAIR principles.
- Statisticians and Analysts: May access pseudonymized data in DataSafe for analysis. They are not granted access to direct identifiers.
- Collaborators (if applicable): Can only access data via secure environments and under data sharing agreements. Their access is restricted based on project needs and reviewed regularly.

3.5 Ensuring data quality is an important part of data stewardship. Please describe your plans with regard to checking and improving data quality. This may be described in a standard SOP from a service or your department, a part of your protocol, part of data processing or built into a data entry system like Castor. Refer to other documents where applicable.

Quality checks (validations) will be implemented in Castor EDC to improve data entry quality. All analyses will be described in (electronic) lab journals and will be cross-checked by the supervisor(s). Data cleaning steps (i.e., check data validity and remove invalid or unwanted observations, adequately code missing data) will be performed. Any changes or manipulations made will be traceable through versioning of documents.
We will discuss within the consortium if we want to write an overall SOP for the quality control of data within the project.

3.6 Do you expect costs for storage and data management during the study or project?

- Yes

3.7 Please describe briefly how these costs will be covered. If you have budgeted for this in your grant application, please specify.

The Healthbox project has budgeted 7.000 euros (in total, so also for trial) for data management.

4. Archiving data within the LUMC

4.1 Which parts of your data will you select for long-term archiving? Please motivate why you would not archive (parts of) your data.

For long-term archiving, only the pseudonymized, cleaned, and processed datasets that are essential for the validation of published results will be selected. This includes:

- Quantitative data (e.g., survey results, eHealth usage metrics, clinical outcome measures) that are used in publications.
- Codebooks and metadata necessary to understand and reuse the archived data.
- Statistical scripts (e.g., SPSS code) used for data analysis.

These data will be archived in a trusted long-term repository in accordance with FAIR principles and institutional policies. Access to the archived data will be determined based on ethical approvals and participant consent (e.g., open or restricted access).
The following data will not be archived for the long term, with motivations:

- Raw data from the HealthBox (e.g., real-time sensor data):

These datasets are extremely large in volume and contain highly granular information that is not required for replicating the research findings. Storing them long-term is not cost-effective and not justified unless they contribute to scientific output.

- Identifiable or personal data (e.g., names, contact information, full medical records):

Due to GDPR and ethical concerns, personal data will not be stored long-term. Only pseudonymized or anonymized versions will be archived, where re-identification is not possible. We will make sure no identifiable data is necessary for reproducibility.

- Administrative or operational documents:

Project planning documents, internal notes, and communication logs will not be archived long-term unless required for audit or legal purposes.

4.2 What will you do to prepare your data for archiving within the LUMC? Describe how you intend to meet LUMC requirements.

1. Data selection

- Only the pseudonymized, cleaned, and relevant datasets used in publications or that support key findings will be archived.
- I will ensure that no directly identifiable personal data are included in the archived datasets, in accordance with GDPR and LUMC data protection policies.

2. Metadata and documentation

- I will create clear and complete metadata following the LUMC guidelines.

Documentation will include:

- Codebooks and variable descriptions
- Study protocol and data collection methods
- File naming conventions and data structure explanations
- Analysis scripts (e.g., R, SPSS)

3. File formats

- I will convert data to sustainable and open formats where possible (e.g., CSV, TXT, PDF/A), as recommended by the LUMC.

- Proprietary formats (e.g., SPSS .sav) will only be included if no suitable alternative exists, and always accompanied by an open version.

4. Archiving platform

- Data will be archived within the LUMC Research Data Repository (e.g., via DataSafe, depending on data sensitivity and access level).
- The data will be stored for at least 15 years, or longer if required by funder or journal policies.

5. Access and governance

- I will define the appropriate access level (open, embargoed, or restricted), based on ethical approvals and participant consent.
- A designated contact person will be listed for future access requests if the data are restricted.

6. Support and quality check

- I will consult the LUMC Data Steward for a final check before archiving, ensuring that:
- The dataset is well-documented and findable
- Personal data are fully removed or anonymized
- The dataset complies with FAIR principles

4.3 Will there be extra costs for this preparation?

- No

4.5 How long must your data be preserved? Please explain briefly in the 'additional information' field.

- Human research WMO/non-WMO: ≥ 15 years

4.6 Are there any requirements regarding the disposal of data?

- Yes

4.7 What are the requirements regarding the disposal of these data?
Describe how you will dispose of the data: how you will get approval, what people and/or tools you need, etc.

Paper informed consent forms will be disposed of in special locked confidential paper containers and will be destroyed as confidential material according to DIN66399-2 guidelines

4.8 How will you ensure data and/or metadata findability and availability within LUMC for the long term?

My department has set up an overview of closed research projects. The department data steward is responsible for archiving data in a dedicated folder. The procedure and responsibilities are outlined in the departmental data stewardship SOP, which can be found in Zenya.

4.9 Do you have costs associated with long-term storage of your data?

- Yes

This depends on the total volume of data when the project is finished. We can't give an estimate yet.

4.10 How will these costs for long-term storage be covered?

It has been budgeted according to our PI

5. Data publication & reuse

5.1 Which parts of your data(sets) will you select for publication?

The following parts of my datasets will be selected for publication, in line with FAIR principles and ethical considerations:

1. Quantitative data

I will publish pseudonymized, aggregated quantitative data that were used in analyses and reported in the results section of publications. This includes:

- FIM (Feasibility of Intervention Measure) scores
- AIM (Acceptability of Intervention Measure) scores
- SUS (System Usability Scale) scores
- Engagement data from the Ancora app (e.g. logins, feature usage, session duration)

These datasets will be cleaned and structured in open formats (e.g. CSV), and will be accompanied by codebooks and analysis scripts to allow reuse and verification.

2. Think-aloud data

- I will publish a thematic coding framework and selected pseudonymized excerpts from the thinkaloud transcripts that support the reported findings.
- Full transcripts will not be made public due to privacy concerns and the potential for indirect identification.
- The methodology and coding scheme will be well-documented to ensure transparency.

3. Focus group data

- I will publish the summarized findings, coded themes, and, where possible, pseudonymized quotations that illustrate key insights from the focus group discussions.

- Complete transcripts will not be shared due to the difficulty of fully anonymizing multi-person qualitative data and the ethical considerations involved.
- Metadata will describe participant characteristics in general terms (e.g. role, age group) to provide context.

5.2 Are there any restrictions placed on sharing/reuse of some/all of your data due to one or more of the following options? Briefly explain the restrictions.

- Intellectual Property (IP; like a patent or licence)
- Funder requirements
- Signed informed consent
- Consortium agreement

The consortium agreement specifies joint ownership of data between the parties. Therefore, all data publication and reuse must be approved by the following institutes:

Universiteit Twente
TU delft
Hogeschool Utrecht
Ancora Health B.V.
Diabetes Fonds

Intellectual property rights are described in the HealthBox consortium agreement, as well as in the already existing agreement between certain participating parties such as the LUMC and Ancora. Additionally they will be discussed with the consortium partners per workpackage.

In our application: Partners are aware of any background existing outside of the consortium likely to constrain the scope of operation planned for the development, demonstration, and exploitation activities. Identifying background before starting the project is essential for preparing the consortium agreement, where the partners will agree on the background of the project. The consortium will work with an actively listed background. Owing to the strong interactions planned between most tasks involving many partners, rules for accessing background knowledge necessary for carrying out these tasks will be defined. For background exchanged in the project and project implementation, the guiding principle shall be royalty-free, full-access basis, unless otherwise determined in the consortium agreement.

When participants invent independently, they can apply for a solo patent after obtaining consent from other participants. If two or more participants jointly invent, they will jointly apply for a patent after obtaining consent from the other participants.

5.3 Will you publish your data open access or with restricted access?

- Partly open and partly restricted access

Partly restricted access because of the think aloud and focus group data. But with a preference for publication in journals with open access.

Goal will be to publish data 'as open as possible, as closed as necessary'. For this we can use the LUMC flowchart that helps us to determine if we have permission to publish a dataset. It will guide us to decide whether we are allowed to publish a dataset online with open access, restricted access or not at all.

5.4 Publishing your data (partly) 'open access': which licence or 'terms of use' will you add?

For software: MIT licence, Apache Licence or one of the GPL Licences
For open access datasets: CC licence or Open Data Commons licences
I will consult with the LUMC Data Steward to ensure appropriate licensing and compliance with institutional and legal requirements.

5.5 Publishing your data (partly) with 'restricted access': what is the reason for this? Please give some background information, if needed, in the 'additional information'.

- Data contains privacy-sensitive information

5.7 Is there an embargo period before publishing your data?

- No

5.9 Where will you publish your (meta)data?

- Publish (parts of) the data in a general database/repository

I will publish pseudonymized quantitative datasets (e.g., FIM, AIM, SUS scores, app engagement metrics) via DataverseNL, a trusted, general-purpose repository supported by LUMC. This ensures persistent identifiers, proper metadata, and long-term findability.
I will not publish full raw datasets or sensitive data in open access repositories.
Qualitative data (think-aloud, focus groups) will only be published in pseudonymized and excerpted form, or under restricted access, due to privacy concerns.

5.10 Give the name(s) and link of the repository, database, website, biobank or catalogue where will you publish your (meta)data and explain what their options are for access.

At the moment I don't know where I will publish my data.

5.11 Usually, the chosen publication format from question 5.10 adds one or more persistent identifiers to your (meta)data. Please specify the type of persistent identifier(s):

DOI

5.12 What will you do to prepare your data and/or metadata for publishing outside the LUMC? Please describe how you intend to meet publisher, funder, database or repository requirements.

Before publishing my data outside LUMC, I'll take the following steps to make sure it meets the requirements of funders, publishers, and the chosen repository:
1. Select and clean the data
I will only publish data that are relevant to the findings I report. All personal identifiers will be removed. For qualitative data, I'll include only coded summaries or selected quotes, not full transcripts.
2. Create clear metadata
I'll describe each dataset using the standard DataverseNL fields (e.g. title, description, methods, variables), so others can understand what the data are about and how they were collected.
3. Use open formats and documentation
Data will be shared in accessible formats (like CSV), with a short readme file and a codebook that explains the variables and structure. If I use any scripts or code, I'll link to them on GitHub/Zenodo.
4. Apply a suitable license
I'll choose a license that fits the level of openness I want—for example, CC BY 4.0 for most datasets, or CC BY-NC 4.0 if the content is more sensitive or meant for non-commercial use only.
5. Check for compliance
Before publishing, I'll consult with the LUMC data steward to make sure everything is in line with privacy rules (GDPR), participant consent, and institutional policies.

5.13 Will there be extra costs for this preparation?

- No

5.15 Do you have costs associated with publication of your data?

- Yes

5.16 How will these costs for publication be covered?

The Healthbox project has set aside a budget for the articles published by Work Package six (WP6).

5.17 Who is responsible for your data and has authority to grant (additional) access to your data after finishing the study or project (e.g. for the long term)?

- Collaborator/research partner organisation
- PI of the study/project
- Data Access Committee (DAC)

Need to figure out if the Healthbox project has a data access committee.

6. Review request

6.1 Please tick the appropriate box.

- Review: feedback and suggestions appreciated