Project Number: *101057390*

Project Acronym: HappyMums

Project title:

Understanding, predicting, and treating depression in pregnancy to improve mothers and offspring mental health outcomes.

D2.1 Data Management Plan, Version 1

Research and Innovation Action

HORIZON-HLTH-2021-STAYHLTH-01-02

Work Package:	2
Due date of deliverable:	30/04/2023
Actual submission date:	05/09/2023
Lead beneficiary:	UB
Contributors:	Catherine Gallin (UB), Carlos Martín Isla (UB), Socayna Jouide El Kaderi UB), Karim Lekadir(UB), with input from all consortium partners
Reviewers:	Libera Cavaliere (UMIL), Annamaria Cattaneo (UMIL).

Disclaimer

Executive Summary

This first version of the *HappyMums* Data Management Plan (DMP) addresses the purpose and description of data handled within the *HappyMums* project and the standards for handling data both during and after the project's life cycle. This includes provisions for data collection, secure long-term storage, integration and interoperability, and accessibility and exploitation, in compliance with the GDPR standards and the principles for findable, accessible, interoperable and reusable (FAIR) research data.

The *HappyMums* DMP is a live document that will be circulated and updated periodically within the project consortium in order to ensure that it is up to date and that any scientific, logistical, or ethical issues related to the use and generation of data are addressed in a timely and efficient manner.

Table of Contents

Acronyms

Abbreviation	Full term
AI	Artificial Intelligence
BAM	Binary alignment map, binary representation of the Sequence Alignment Map-files
CSIR	Chronic social isolation rearing
CSV	Comma separated values
CUDA	Compute Unified Device Architecture
DMP	Data management plan
DPA	Data Processing Agreement
DPO	Data Protection Officer
DTA	Data Transfer Agreement
EGA	European Genome phenome Archive
ELS	Early -life Stress
EU	European Union
FAIR	Findable, Accessible, Interoperable, Re-usable
FCSIR	Fish chronic social isolation rearing
FPNS	Fish prenatal stress
FL	Federated Learning
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
MSUS	Unpredictable maternal separation combined with unpredictable maternal stress
PNS	Prenatal Stress
RAM	Random Access Memory
WP	Work Package

1 Introduction

Perinatal mental disorders contribute enormously to societal and health burdens, both during pregnancy and beyond. According to the World Health Organization (WHO) data, depression is the most common mental health disorder in pregnancy, affecting between 10% and 15% of pregnant women (1). In turn, a mother's mental health and wellbeing may affect their child's development. Stress, depression, and anxiety during and after pregnancy are associated with disturbances in childhood development.

HappyMums will be the first EU project specifically designed to improve our understanding of the biological mechanisms underlying the development of depressive symptoms in pregnancy, and the efficacy of interventions. *HappyMums* will also identify the mechanisms that affect fetal environmental biology, shaping offspring's risk for developing negative mental outcomes later in life. Further, we will identify the prenatal and postnatal factors that exacerbate, or buffer, the risk shaped in utero. The *HappyMums* consortium will also develop a digital platform where different biomarkers (AI tools-based data integrated with biological, clinical, medical, environmental, and lifestyle data) will be collected through a mobile phone app that will be at the interface with clinicians. This will allow early screening of depressive symptoms, prompt diagnoses, personalized treatments, and the promotion of protective lifestyle attitudes. Importantly, by improving mental health in mothers, *HappyMums* will provide unpreceded benefits also to the offspring and thus to society at large.

At the project's core are the following six main objectives:

1. To **identify risk and protective factors for the development of depressive symptoms in pregnancy** and associated biological blood signatures (WP2, WP3).
2. To **characterize the impact of prenatal maternal depressive symptoms on offspring** mental outcomes across development and identify underlying biological mechanisms as well as pre- and postnatal moderators (WP2, WP4).
3. Using animal models, to **identify peripheral and brain molecular mal-adaptations associated with depressive symptoms during gestation,** and dissect the role of placenta in mediating negative mental outcomes development in offspring (WP6).
4. To **develop an innovative digital platform for an app-based early screening, diagnosis, and personalised management of maternal depressive symptoms** in pregnancy (WP5, WP8).

5. To **integrate human and animal data to test novel interventions** to prevent or to improve depressive symptoms during pregnancy and to dissect the underlying biological mechanisms in mothers and in the fetal environment (WP7).
6. To **develop actions to improve wellbeing in pregnancy, to update clinical practices, and to promote the uptake of digital technologies** (WP9).

In order to meet these objectives, *HappyMums* will leverage and generate various types of data, including clinical data from hospitals, data collected from digital tools, behavioral, imagining, and biosample data from animal models, various types of data from longitudinal cohort studies, and public video datasets for facial and emotion recognition. Please refer to Section 3 and Annex I of this DMP for more details on the type and purpose of data involved in the project.

2 Overview of Data Management Plan

The purpose of this Data Management Plan (DMP) is to outline and detail the purpose of the use of data and to detail the data which will be used in the *HappyMums* project, as well as how this data will be handled both during the course of the project and after the project ends. This document therefore provides an overview of:

1. The type of data that will be collected and analyzed, including details on which institutions and work packages are involved with each dataset.
2. How the data will be stored.
3. Ethical considerations concerning the project data and potential mitigation measures.
4. How the use of data will follow the principles of findability, accessibility, interoperability, and reusability (FAIR).
5. How the use of data will follow the European General Data Protection Regulation (GDPR).
6. How the data will be handled after the end of the project.

The *HappyMums* DMP is a live document, which will be regularly revisited and updated throughout the project's life cycle. To this end, regular meetings will be scheduled, and data management registers will be circulated to ensure that information from all partners is kept up to date and that any potential risks or issues are detected and mitigated in a timely manner.

3 Data Summary and Purpose

HappyMums will leverage diverse large-scale datasets in order to explore the physiological mechanisms of perinatal depression to enable effective science-based, clinical interventions to support endangered mothers and improve fetal environment and development.

Laboratory research models will be juxtaposed with unique human samples of the placenta, chorionic villi, and amniotic fluid to unveil the root biological causes of maternal depression. A digital platform will offer rich medical data and a mobile phone app to support mothers and their children.

Fifteen of the consortium's seventeen partners will be working directly with data, across a wide range of data types, coming from both human studies and animal models, in order to achieve the five key concepts of the *HappyMums* project, as shown in Figure 1

Concept 1: Multimodal predictors for depressive symptoms, their progression, and predictors for response to interventions.

Concept 2: Biological signatures and markers linking maternal depressive symptoms to offspring outcomesand assessment of the role of postnatal factors as moderators.

Concept 3: Neurobiological mechanisms and peripheral biomarkers by using preclinical models.

Concept 4: Development of personalized strategies to improve mental health in pregnant women by usingmachine learning approaches and digital tools.

Concept 5: Effect of pharmacological and non-pharmacological interventions in pregnancy and underlying biology.

Figure 1: Key concepts of the HappyMums project

The data used and generated in *HappyMums* will be useful to the following groups:

1. Women who are pregnant or planning to have a baby.
2. Researchers working in the field of depression in pregnancy.
3. Healthcare professionals (nurses, midwives, clinicians, perinatal psychologist and psychiatrists, etc.
4. Clinical researchers interested in biomarkers for a better diagnosis and mental health monitoring.
5. Public bodies and regulatory agencies working on preventive policies, clinical guidelines, and funding allocations.
6. Private and public investors interested in developing innovative tools for screening, early risk identification and interventions in pregnant women.

3.1 Types of Data

In *HappyMums,* quantitative and qualitative experimental data from humans (combination of existing and newly generated data) and animal models (newly generated data) will be collected. In particular,

- Molecular numerical data (DNA, RNA, epigenetics, metabolomics) from human cohorts' studies (both existing and newly generated) (WP3, WP4, WP7) and rodent experimental models (WP6, WP7).
- Clinical and behavioural numerical data in humans (combination of existing and newly generated, WP3, WP4, WP7) and behavioural numerical data in animal models (WP6, WP7).
- Bioinformatical data (numerical): pathway analysis and integration analysis of molecular and behavioural data in humans and animals (WP3, WP4, WP6 and WP7).
- Neuroimaging data (images, numerical) from existing human studies (WP4) and generation of Neuroimaging data (images, numerical) from experimental animal models (WP6, WP7).
- AI quantitative data and images from the *HappyMums* mobile phone app (WP5, WP8).

HappyMums will reuse existing data as well as generate new data, as shown in detail in the data inventory table in Annex I.

For a brief summary of datasets to be used in the project, please see Table 1 below.

Dataset	Data Type	Sample Size
Berlin Birth Cohort Study	Longitudinal cohort study	N = 362
PREDO	Longitudinal cohort study	N = 4,777
ITU	Longitudinal cohort study	N = 943
PRAM-D	Longitudinal cohort study	N = 172
Imprint	Clinical intervention study,	N = 1000
PRESeNT	Clinical intervention study,	N=200
GenerationR	Longitudinal cohort study	N = approx. 10 000
GenerationR Next	Longitudinal cohort study	N = *approximately* 5,000 (*approximately*: N = 1,500 preconception; N=3,500 during pregnancy)
Brazilian High Risk Cohort Study (BHRCS)	Longitudinal cohort study	Phase 0 (from age 6): N = 712 Phase 1 (from age 9): N = 447 Phase 2 (from age 12): N= 404

EU Child Cohort Network	Network of cohort studies started in early life	N = 250,000
Public clinical registers	Clinical data, longitudinal	N = approx. 800,000 500,000 from SLaM (London), + 300,000 (Italian hospitals), exact number to be confirmed.
HappyMums Multi Center Clinical Study	Digital data from *HappyMums* clinical study	N = 1, 000 However, no data collected yet. Data will be collected during the *HappyMums* clinical study.
AffectNet	Video	N = 400,000
D-vlog: Multimodal Vlog Dataset for Depression Detection	Video	N = 961
Audiovisual Emotion Challenge (AVEC)	Video	N = 292
Pittsburgh Depression Dataset	Video	N = 57
Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D)	Video	N = 7,442
Sewa Database	Video	N = 180
UMIL Mouse model	Animal experiments, behavioural, imaging, biosamples	N=30-40 per group for behavior, n=10-15 per group for imaging and biosamples
UZH Mouse models	Animal experiments, behavioural, imaging, biosamples	N=30-40 per group for behavior, n=10-15 per group for imaging and biosamples
Cognitive flexibility training in mice (COLUMBIA)	Animal experiments	N=10-15 mice per group
Animal model of PNS/ELS (UvA)	Animal experiments, microglia profiling of dams and offspring	N=16 dams N=16 male and 16 female offspring. 8 for ex-vivo functional assay and 8 for morphological characterization
Animal models of ELS (PNS/ELS, MSUS, CSIR) ex vivo MR imaging (KCL)	Animal experiments, neuroimaging	WP6: N=72 mouse dams; N=288 mouse offspring. WP7: N=128 offspring.
Livebearing fish (FCSIR and FPNS)	Longitudinal cohort study	N=500

Table 1: Summary of data used in HappyMums

In general, data and metadata will be gathered, stored, and transferred in a comma-separated values (CSV) format or tab- delimited text (.txt). In addition, MS Excel

compatible files will be included to facilitate the collected data exchange. Raw data such as BAM files will be generated from e.g. RNA sequencing analysis. For more details on storage plans per each consortium partner, please refer to the detailed data inventory in Annex I.

3.2 *HappyMums* Clinical Study

As stipulated in the *HappyMums* Grant Agreement, clinical criteria and study design will be agreed within the first 12 months of the project by clinicians withing *HappyMums* and patients' associations (T5.4). Discussions are currently underway between the clinical partners, with periodic clinical data meetings and continuous contact to discuss which specific biomarkers should and can be collected. The study definition is not yet complete, with the aim of completing the parameters and methodology set for the end of 2023.

3.2.1 HappyMums multi centres clinical study

Once developed (M12), the *HappyMums* app will be used for a large screening across West and East Europe on a population of pregnant women at high risk for depression or already suffering from depressive symptoms. This clinical study is to be conducted across 7 clinical sites. The master protocol, although standard, will be produced by KCL, while individual handling of site data will be dependent on the laws and ethical approvals at each site. The exact details of data sharing between ABACUS, the partner in charge of developing the app, and other sites, are yet to be confirmed. This will be decided when ethical applications are submitted and approved, around M12-16. Preliminary details can be found below.

The *HappyMums* mobile app that will be implemented in the initial phases of the project will be used in the *HappyMums* multi centres clinical study. It will be used by each woman involved in the study and it will generate data from different sources:

- **Digital version of clinical questionnaires** (e.g. GAD7, PHQ8, EPDS) **and tests** (e.g. Infant Emotions Recognition) presented in the app at predefined time intervals.
- **Game-like activities** presented in the app and spontaneously executed by the woman (e.g. Mood Tracking, Describe your day, Spatial Memory). Raw data collected from game-like exercises will be of different types: audio files, videos, text files.
- **Completion level of Mental well-being course** given inside the app at each time the course is accessed.
- **Smartphone background sensors data** like GPS, accelerometer, phone usage passively collected in the background.

- **Spontaneous manual data entry** for monitoring purposes (e.g. weight, blood pressure).
- All data generated by the use of *HappyMums* mobile app will be securely transferred, stored and processed in a protected cloud server located in Europe complying with data security and ethical standards of *HappyMums* project described in section **Error! Reference source not found. Error! Reference source not found.**.

4 Implementation of FAIR Principles

The FAIR principles (**F**indability, **A**ccessibility, **I**nteroperability, and **R**euse) of data and digital assets are at the core of the *HappyMums* project. This DMP will be periodically reviewed and updated to ensure the continual application of FAIR principles across the whole project and all consortium members.

HappyMums will implement the FAIR principles by making the collected data accessible in the following platforms:

1. A dedicated *HappyMums* Research Platform (WP2), designed by UB.
2. The EarlyCause Data Portal, as described in Section 4.2 below.

All institutions will provide relevant metadata and keywords to have the collected data easily discoverable. Standard naming conventions will be used according to each type of experiment. Keywords will be provided to easily re-use the generated data. All version numbers provided will be clear and up to date.

The *HappyMums* Research Platform (WP2) will use the formal, standard, accessible, shared, and broadly applicable RDF language for representation and include qualified references to other data. We will use standard coding schemes for disease and medications, such as ICD10, ATC, and RxNorm. Data and metadata vocabularies, standards, or methodologies used to make data interoperable are standard vocabularies. Reusability of data/research outputs: if some data will be licensed to permit their re-use by the external stakeholders, Creative Commons or GNU licenses will be used. When making e.g. newly generated omics and (pre)clinical data public, we will do so with incorporating journal requirements regarding embargo periods around publications.

4.1 Metadata and Data Harmonisation

4.1.1 Overview of use of metadata

Metadata will have a two-fold nature: descriptive, giving information on the data discovery and identification (titles, author, keywords) and administrative outlining when and how data was created, file type and other details, and who can access it. Compute plans in federated analysis will be archived and their execution will be logged with secure authentication. Accessibility of data/research outputs: *HappyMums* plans to use the multiple external deposit research data which support open access and are relevant for our disciplinary approaches, such as LIPID MAPS, Gene Expression Associated with document Ref. Ares(2022)7741729 - 09/11/2022 101057390 — *HappyMums* Part B 25 Omnibus/GEO, PeptideAtlas (some partners already have experience with depositing collected data and metadata on GEO). Additionally, we will establish a dedicated

HappyMums Research Platform (WP2) and use institutional repositories to ensure Open Access to our data following the principle "as open as possible and as closed as necessary", with appropriate licenses. We will specify the conditions for the access, which will allow full download and semantic querying using SPARQL (SPARQL is available in the free R and Python languages).

Metadata curation and harmonisation will be performed early in the project to allow interoperability across cohorts. Protocols to harmonise longitudinal cohorts will be applied to cover: i) Presence of depressive symptoms, their severity, and diagnoses, ii) Biological data iii) Lifestyle, psychosocial and environmental factors. Biological data will be integrated including a specific harmonization pipeline to reduce the impact of between and within centres batch effects, using ComBat algorithm in combination with visualization checks (e.g., PCA plots, dendrograms, or heat-maps).

4.1.2 Overview of data harmonisation

Metadata curation and harmonisation will be performed to allow interoperability across cohorts. Protocols to harmonise longitudinal cohorts will be applied to cover: i) presence of depressive symptoms, their severity, and diagnoses, ii) biological data iii) lifestyle, psychosocial and environmental factors. Biological data will be integrated including a specific harmonisation pipeline to reduce the impact between and within centres batch effects, using ComBat algorithm in combination with visualization checks (e.g., PCA plots, dendrograms, or heat-maps). Discussion on centralisation of biological samples and other data is ongoing and will be defined in the coming months.

The details on harmonisation processes are currently being discussed in periodic work package meetings and can be categorized as follows:

1. **Hospital Clinical Data.** KCL and OSR will work together (T3.1) to analyse data from clinical registers (SLaM of up to 500,000 patients in UK, and a population-based registry, known as Regional Data Warehouse (DWH), in Italy (Lombardia). These data are fully anonymised, but specific application will need to be made to perform research on these data. We are now in the process of preparing specific project application taking into account also the challenges that the use of clinical register can present, including data inaccuracy, incompleteness, and biases, along with ethical issues that can limit the potential of data integration. Data on the association between available risk or protective factors (e.g presence or absence of other pathologies, previous history of depression or other psychiatric conditions), and clinically relevant outcomes (e.g needs for care for peripartum

depression, discharge from services, and hospitalisation), obtained through penalized regressions, will be available by M18 (30 April 2024).

2. **Pre-existing datasets (longitudinal cohorts, etc.)** Data and samples (cord blood samples, placenta, amniotic fluid, blood samples from women, saliva from children) collected in ongoing medical studies carried out in different European countries (Germany, UK, Finland, Netherlands) and Brazil will be exploited in WP3 and WP4. The harmonization procedures will involve the following steps: identifying data domains to be harmonized, compiling an inventory of relevant data across cohorts, defining project-specific research protocols and selecting participating studies, harmonizing study-specific variables under a common format, estimating the quality of the harmonized data, preserving the harmonized data and related metadata for reproducibility and future use. This will be done both for clinical and biological variables. Specific Standard Operating Procedure in this regard are being developed, under the supervision of OSR, and will be described in D2.3. *Data Harmonization Procedures and Protocols* due at M10.
3. **Mobile app data** that will be collected within *HappyMums* in T5.4 Screening of pregnant women in a multi centrers clinical study. The harmonisation will pertain:
    a. **Questionnaires, cognitive tests, etc.:** the same questionaries will be used across the different centers, using the appropriate validated translated versions. Data will either be stored centrally by ABACUS, or on site-specific servers e.g. RedCap
    b. **Video/audio data** video pre-processing will be done to harmonise such videos in resolution/ framerate /clip extraction/ brightness/contrast and facial ROI extraction, due to the high variability in mobile optical sensors. For Audio data, sample rate and amplitude will be normalised across devices. Responsive interactive assessment will be included in-app to guide the user along the recording.

4.2 Data Submission to the EarlyCause Data Portal

For all partners who wish to and whose local legislation permits, various data types used in *HappyMums* will be deposited into the relevant archives within the EMBL-EBI infrastructure through the EarlyCause Data Portal (https://earlycause.eu/, Horizon 2020 Project Grant 848158). EarlyCause is a Horizon 2020 EU project (grant number 848158) coordinated by *HappyMums* partner UB. The EarlyCause Data Portal was created by and is maintained by EarlyCause partner EMBL-EBI. The well-established EMBL-EBI infrastructure, consisting of a network of trusted archives which are widely used

throughout the scientific community, therefore provides the solid basis for the EarlyCause Data Portal as well as guaranteeing continued accessibility beyond the project life cycle.

The EarlyCause portal represents the front-end access point to the ELS-related data within the underlying archives. Since Open Science and FAIR data sharing principles have always been core values of EMBL-EBI processes, the policies and infrastructure of the portal have been designed with Open Science/FAIR principles in mind. Data is being shared as openly as possible and as closed as necessary, while the archives also provide appropriate data models to capture metadata that facilitate FAIR principles.

The EarlyCause Data Portal utilises the existing software framework which is also being used in the European COVID-19 Data Portal and the Pathogens Portal; this contributes to the interoperability and sustainability of the platform and has also initiated a collaboration with the Molgenis cohort catalogue. This collaboration addresses the important aspect of appropriate data standards, data structure and data models to ensure the **interoperability** across multiple platforms within the same field, to maximise **findability, accessibility**, and **re-useability**, in line with FAIR principles. Internal discussions were held within EMBL-EBI to further embed the system into the existing infrastructure and a long-term strategy is currently being implemented.

The first functional version of the EarlyCause Data Portal sits on the project website and can be accessed via https://portal.earlycause.eu/, where it is currently available to the wider research community and the general public, although it continues to be a work in progress. The portal allows the user to browse early life stress related datasets and offers search and filtering functionalities. Further improvements and optimisations will continue to be implemented throughout the EarlyCause project's lifecycle and the data portal will remain active after the EarlyCause project end within the EMBL-EBI and ELIXIR framework.

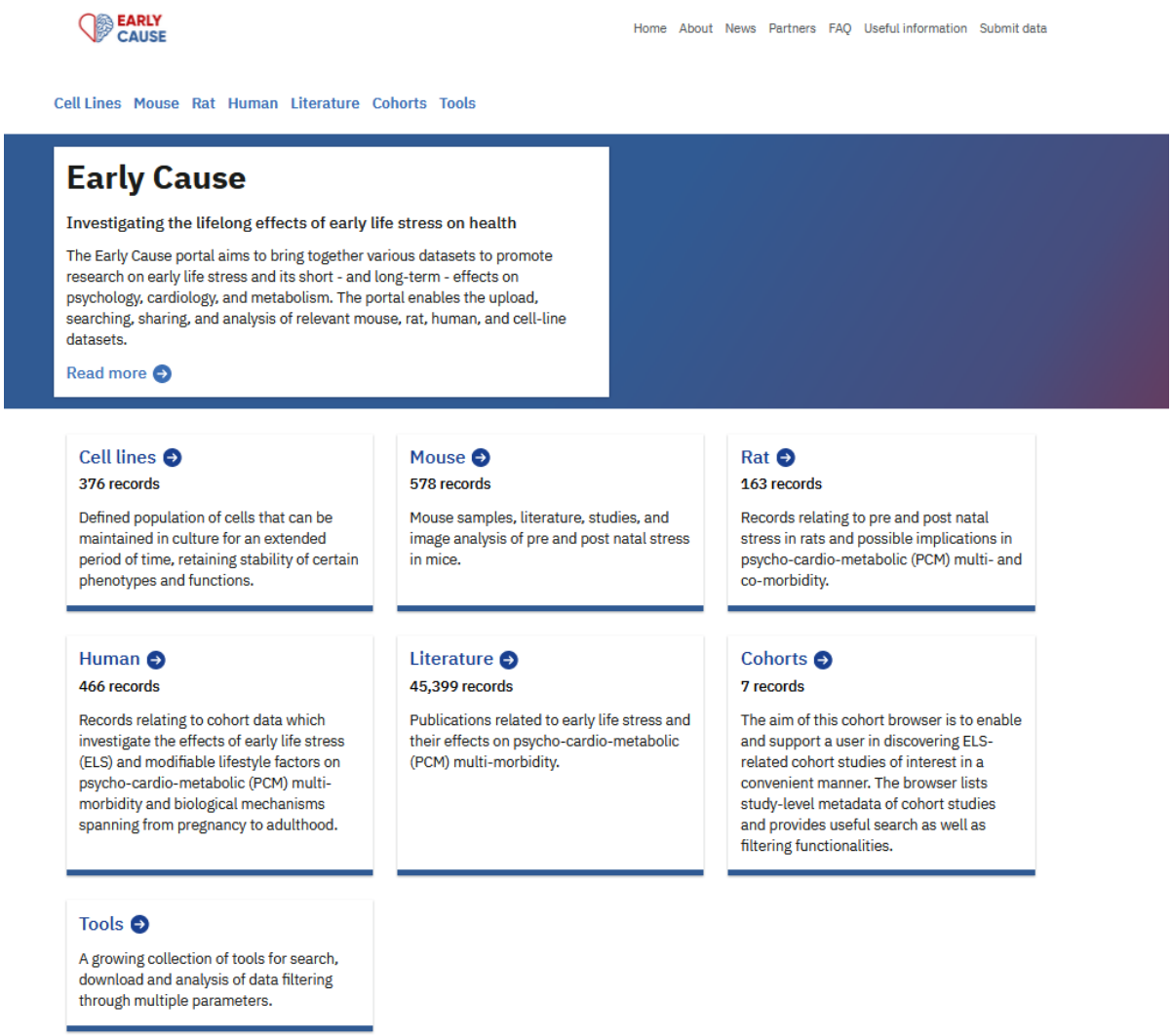

Figure 2: Screenshot of the EarlyCause Data Portal, https://portal.earlycause.eu/

The portal currently consists of several components and will include project protocols for further explorations in the future. Relevant datasets can be found via various sections shown in the screenshot below. Each section contains project accession numbers generated from their respective repositories at the time of submissions. The list of projects can also be filtered using data types, lab affiliation etc.

When submitted, data from the *HappyMums* project will be assigned a "*HappyMums*" keyword so that datasets from our project can be clearly identified amongst the datasets from EarlyCause and other projects. Within the EarlyCause Data Portal, a project can be searched directly using the 'Search' field if the permanent identifiers are readily available on hand so that they can be presented in the EarlyCause Data Portal which serves as a central point of access for ELS-related data (from EarlyCause and other publicly available datasets).

4.2.1 Categorization of data subtypes on portal

Data types will be deposited into their corresponding archive on the EarlyCause Data Portal, examples of which are shown in Table 2 below. As discussions within *HappyMums* continue, we will outline which partners will be contributing which data types and their corresponding database; this table will be updated accordingly as needed.

Database	Data Type
European Nucleotide Archive (ENA)	Molecular sequence data from animal models
BioSamples	ELS exposure and mouse model stress descriptors
BioStudies	Mouse behavioural data
MetaboLights	Metabolic profiling data
BioImage Archive	Image data
PRIDE	Proteomics data
ArrayExpress	Gene Expression data
European Genome-phenome Archive (EGA)	Personally identifiable genetic and phenotypic data resulting from biomedical research projects

Table 2: Submission of data type by archive, EarlyCause Data Portal

5 Data Security and Ethics

The *HappyMums* project will be compliant with the European Union (EU) 2016/679 General Data Protection Regulation (GDPR). Provisions for data security include best practices to guarantee data security and logical data security (i.e. using firewalls, intrusion detection system, authentication methods, logging of activities on file systems where data is stored). The data recoverability is granted by backup systems based on best-of-breed enterprise backup software as well as enterprise level backed-up data repository (virtual libraries and hardware repositories with data deduplication technology enabled). We will leverage the EarlyCause Data Portal (see above Section 4.2) and will investigate connection to the European Health Data Space and European Open Science Cloud, and especially to federated extensions of Elixir (Federated Human Data).

The data inventory (Annex I) includes potential ethical risks and corresponding mitigation measures as well as the details of the ethical approvals of each partner institution. The ethical and security issues of data management will be continuously reviewed throughout the project's life cycle both through regular DMP updates as well as during consortium meetings and through the project's critical risk register.

5.1 GDPR Compliance Principles

Ensuring GDPR compliance principles is an ongoing effort in the *HappyMums* project unfolding on three levels in terms of creating data transfer and processing agreements governing data transactions and translating legal and ethical requirements into technical ones. More specifically the following actions, tools and measures will be implemented to ensure compliance with GDPR principles (art. 5). Respectively:

- **Lawfulness, fairness and transparency:** Data access procedures, implemented in each of the storage repositories will include the collection of justification for data use. Each data provider will monitor the fulfilment of this requirement, including guarantees stated in art. 89 (as pseudonymisation and revision by local ethics committee) under the guidance of their local ethics and legal committees. The imaging and non-imaging data will enforce data access procedures, in accordance with GDPR principles via business rules.
- **Purpose limitation:** The overall purpose for data use in *HappyMums* is scientific research and so it will be stated in all DTAs, DPAs, policy documents and legally binding disclaimers clicked-through by data users on the *HappyMums* Research Platform. Such purpose is in line with core project's objectives and recognised by both the GDPR and each participating member state law. Each data provider (data

controller) will implement this particular requirement (as per art. 6.4 and 9.2 j) or alternatively ask for data subject explicit consent for other purposes, where applicable. No secondary uses will be allowed without explicit consent as indicated in the DTAs, DPAs and disclaimers.

- **Data minimisation:** Clinical and AI researchers are already actively applying this principle in the definition of clinical requirements with the selection of data strictly needed for the AI/clinical tasks. In current activities, such definitions are implemented in technical annexes to DTAs and DPAs in relation to the use cases.
- **Accuracy, storage limitation, integrity and confidentiality**: Storage limitations in *HappyMums* will be reviewed, and extensions proposed if needed to align these with the project's policies developed. These revisions will focus on terms regarding retention period, implementation of right to be forgotten and other subjects rights, related policies, tools and procedures.
- **Accountability** will be enforced primarily with the indication of data controllers, and type of controllership, at each step of the data flow. Respective national authorities under which such roles and their responsibilities will be enforced, will be stated in DTAs and DPAs. Ethical reviews have either been approved, are in the application process, or are being prepared, as detailed in the data inventory in Annex I.
- **Data security:** Provisions for data security include best practices to guarantee data security and logical data security (i.e. using firewalls, intrusion detection system, authentication methods, logging of activities on file systems where data is stored). The data recoverability is granted by backup systems based on best-of-breed enterprise backup software as well as enterprise level backed-up data repository (virtual libraries and hardware repositories with data deduplication technology enabled). We will investigate connection to the European Health Data Space and European Open Science Cloud, and especially to federated extensions of Elixir (Federated Human Data).

6 Inter-Partner and Cross-Border Data Sharing

HappyMums will leverage several large-scale phenotype-rich studies in Europe and North and South America. Cross-border data sharing is being carefully examined, also in cases of data sharing and accessibility within different European countries. Consequently, adequate data handling and management is crucially important, which we tackle by the adoption of novel technologies, such as federated data analysis and linked open data (LOD) based on collaborative repository.

In all cases, GDPR standards and as well as local standards concerning cross border data access, adequacy decisions to be taken into account, de-identification measures, standards and terminology applied in EU and third countries will be addressed.

The local legislations of each consortium partner concerning the deposition of data in fully open and controlled data access repositories will be examined throughout the project. These issues and other potential data sharing and data access standards will be discussed during regular data management meetings and correspondence between the project's data management team and the consortium partners throughout the project; regular data management registers will also be circulated amongst all partners in order to keep track of any emerging legal, ethical, or scientific issues.

6.1 Federated Learning Platform

Federated learning will allow large-scale data integration while ensuring institutional and personal privacy by limiting data integration within the scope of the applied models and preserving further characteristics of the data for its respective owner. A collaborative semantic repository using linked open data will allow the sharing of data, knowledge, and summary statistics with an intermediate status between to support efficient reuse of expertise, data, and computation. For this purpose, the *HappyMums* project will utilize the Federated Learning System of the DataTools4Heart (DT4H) Horizon Europe project (Grant Number 101057849), coordinated by the UB.

The high-level architecture of the DT4H federated processing system involves a central platform that coordinates the collaboration between different centres (or nodes). The platform is built using state-of-the-art technologies that enable secure data sharing and processing, while ensuring the privacy and confidentiality of the data.

As part of our Federated Processing system, the project has developed a distributed analysis engine that can deploy and support the entire life cycle of a Federated Learning (FL) experiment or other federated analysis.

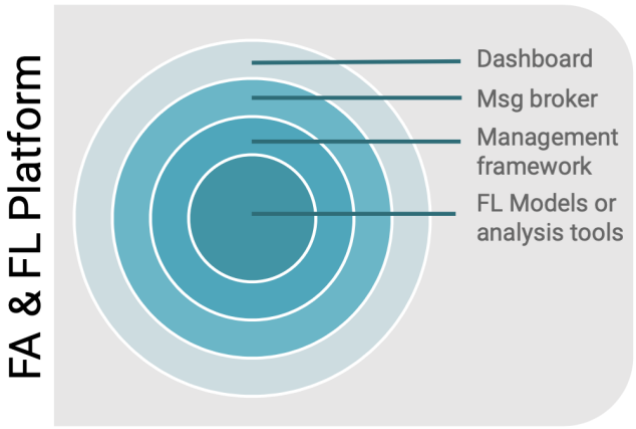

Figure 3: Federated learning use-case architecture overview

As depicted in Figure 3, for a federated learning use-case we would have the following architecture:

- Dashboard: user interface where the user triggers an experiment and retrieves the results.
- Message broker: with which we establish the communication between the nodes and central node.
- Management framework: In order to trigger the experiment at each node, a docker-based local implementation of the flower client, at the local node, and flower server, at the master node, are instantiated.
- Federated learning models or analysis tools: these are materialised using a GitHub and mounted into the flower client docker image.

Nevertheless, this would be adjusted based on the analysis that is going to be performed instead of federated learning.

1. Non-functional requirements:
2. Hardware: GPUs that support CUDA and have *at least* 12GBs of VRAM and 32GBs RAM, recommended: GPUs with more 24GBs VRAM, 64Gbs RAM.
3. Allowed outgoing network connection to be able to connect to the central services of the FL platform.
4. Allowed software installation (docker or singularity containers).

Further discussions and decisions on the implementation of the federated learning platform as well as the storage of all clinical data will be updated throughout the project. In particular, we expect more concrete decisions to be made about the local storage of analytes at the end of 2023, when local ethical reviews are expected to have been submitted at partner institutions.

7 Allocation of Resources

Aspect	Response/explanation
What will the costs be for making data or other research outputs FAIR in *HappyMums* (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)?	While the metadata about the datasets will be maintained and served by Metadata Catalogue, the actual data set will be kept in a generic format, such as CSV, in hard drives and virtual drives, depending on the case and the partner. Although, the overall cost would be in proportion to the data size, data to be collected/generated in *HappyMums* is envisioned to in scale of 10GBs, which would not incur a substantial cost.
How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)	The budget of *HappyMums* will cover such costs if compliant with the conditions described in the Grant Agreement.
Who will be responsible for data management in *HappyMums*?	UB is the WP2 leader and will make sure to update the Data Management Plan considering the input of the technical partners involved.
How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?	To be defined.

Table 3: Allocation of resources, data management HappyMums

8 Data Protection Officers

The monitoring of the different issues related to data management is an integral part of the *HappyMums* project, for which reason we have gathered the contacts of the data protection officers (DPOs), which can be seen in Table 4, below. The UB, as well as partners OSR and UMIL, will coordinate regular meetings and/or correspondence with the clinical data partners and DPOs will be notified of any relevant issues, including pertinent issues that need to be solved related to data transfer agreements (DTAs) and ethical approvals.

Partner	Name	Role	Email
UMIL	Pierluigi Perri	Professor on Data Protection & Data Governance.	pierluigi.perri@unimi.it
EMC	Data Protection Office	DPO	functionaris.gegevensbescherming@erasmusmc.nl.
OSR	Data Protection Officer	DPO	dpo@hsr.it
UB	Ruben Ortiz	DPO	protecciodedades@ub.edu
AU	To be assigned Rodrigo Grassi-Oliveira	DPO Associate Professor	N/A rogo@clin.au.dk
ABACUS	To be appointed if needed	DPO	N/A
UH	Petri Hämäläinen	DPO	eutietosuoja@hus.fi
CHARITE	Data Protection Office	DPO	datenschutzbeauftragte@charite.de
MARCE	N/A	N/A	N/A
CUC	Martina Tolić	DPO	martina.tolic@unicath.hr
UvA	Barbara Gerretsen LLM	DPO	fg@uva.nl
WUR	Gerda Bakker	DPO	gerda.bakker@wur.nl
COLUMBIA	COLUMBIA	Emechete Ejike	Chief Info Security Officer
SWPS	Piotr Liwszic	DPO	pliwszic@swps.edu.pl
UZH	UZH Data Protection Office	DPO	privacy@dsd.uzh.ch
KCL (animal data)	Information Compliance	DPO	info-compliance@kcl.ac.uk
KCL (clinical data)	Kristi Sawyer (tentatively)	Senior postdoctoral researcher	kristi.m.sawyer@kcl.ac.uk
TOMMY'S	N/A	N/A	N/A

Table 4: List of HappyMums DPOs

9 DMP Roadmap: Next Steps

In this roadmap (Table 5), we distinguish between internal and external releases/updates of the DMP. External releases will be linked to the project's deliverables, while internal releases will be made regularly to reflect changes and updates in data, technologies and/or policies. Internal releases will be available to the project partners at all times and, if relevant and appropriate, may be released to outside parties on request.

In addition, in consortium meetings, this first draft of the DMP will be discussed as well as whether there is a need for a data management/data access inter-partner committee.

Date	Internal update/ external release	Description
July 2023	Version 1	Deliverable D2.1, DMP version 1, PU
December 2023	Updated version 1.1	This internal version will include an update on the status of the *HappyMums* Clinical study and the *HappyMums* digital platform, as well as any updates on data harmonisation procedures and ethical approvals.
June 2024	Updated version 1.2	We expect two internal DMP releases a year, one in June-July and one in December-January, to reflect the changes that can be triggered by a number of events, such as the modification of clinical use cases, new collaborations with other clinical centres, specific updates or additions to the data model, technological advancements, changes to laws and regulations that affect topics within the scope of the DMP, or new insights and standards that emerged from project progress and inter-project collaborations.
December 2024	Updated version 1.3	
June 2025	Updated version 1.4	
October 2025	Version 2	Deliverable D2.5, DMP version 2, PU
January 2026	Version 2.1	Periodic internal updates to the DMP
June 2026	Version 2.2	
October 2026	Version 2.3	

Table 5 Data management roadmap

10 Conclusion

The *HappyMums* DMP, which provides an overview of the type, purpose, storage, and ethical use of data throughout and beyond the project's lifecycle, is a live document that will be updated periodically during the project, led by the UB partner but in continual collaboration with all consortium partners working with data. These updates will ensure the smooth functioning of data analysis across all work packages and tasks, but also to monitor the ethical implications of data use and compliance with the GDPR and the FAIR principles.

11 References

1. WHO, Mental Health and Substance Use, Maternal Mental Health, https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/maternal-mental-health
2. EarlyCause Data Portal, https://portal.earlycause.eu/

Annex I: Data inventory, storage, and ethics

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
HappyMums completed studies: Already existing cohorts, where the recruitment has been already completed, biological samples already collected and some of the biological data needed in *HappyMums* already obtained.									
CHARITE	Berlin Birth Cohort Study	Germany	Data will be stored on Charité secure server, only available to those with approved access.	Longitudinal cohort study: imaging data, clinical data, bio-samples	Recruitment already completed; Biological Samples collected. Some of the biological data have already been generated; others will be generated in the next months.	Clinical data, biological samples and already generated data will be used in *HappyMums* implementation (e.g WP2,3,4).	The Berlin Birth Cohort Study (CHARITE) includes data from three different studies, two pregnancy studies (Meks study and TransCT study) and one infant study (K2H study). For participants from all three studies: stored (min. 10 years) on electronic data carriers and evaluated in pseu-donymised form. Data can be only shared with cooperation partners who are mentioned in the ethics documents. For participants, who signed the 5th K2H Amendment (10.03.2022) for data sharing: data from all 3studies (antenatal and postnatal) can be shared with new cooperation partners (who must be mentioned on the website)	For participants, who signed the 5th K2H Amendment (10.03.2022): data from all three studies (antenatal and postnatal) can be shared with cooperation partners, mentioning of cooperation partners on the website. For data of all other participants with missing consent of the 5th K2H Amendment (10.03.2022) for data sharing : currently evaluating data sharing options with technical transfer team.	Ethics documents from the Meks study are approved by Charité-Universitätsmedizin Berlin and include: Original ethics documents, approved 01.07.2016 1. Amendment 08.05.2018 2. Amendment 15.04.2020 Ethics documents from the TransCT study are approved by Charité-Universitätsmedizin Berlin and include: Original ethics documents, approved 08.07.2019 1. Amendment 05/24.03.2020 2. Amendment 04.08.2020 3. Amendment 11.05.2022 Ethics documents from the K2H study are approved by Ludwig-Maximilians-Universität München (study has been carried out in cooperation with Ludwig-Maximilians-Universität München) and include: Original ethics documents, approved 17.10.2018. 1. Amendment 14.02.2019

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
									2. Amendment 17.03.2020 3. Amendment 29.06.2020 4. Amendment 15.09.2020 5. Amendment 10.03.2022 (data sharing amendment)
KCL	Psychiatry Research and Motherhood - Depression (PRAM-D)	UK	Data will be stored on KCL secure server, only available to those with approved access.	Longitudinal cohort study with collection of clinical data and biosamples.	Recruitment already completed; Biological Samples collected. Some of the biological data are generated.	Clinical data, biological samples and already generated data will be used in *HappyMums* implementation (e.g WP2,3,4).	Data are pseudo-anonymised, linked to an ID number which cannot be linked to an individual's name. Any personal identifiers are stored separately.	Data sharing would need to be approved by NHS and/or KCL ethics committees, and may involve re-approaching participants to gain consent. KCL suggest that a harmonised data analysis plan is created, then analyses will be run-in house, and summary statistics shared.	07/Q0703/48 (22/05/07) Most recent amendment 30/06/12
UH	Prediction and prevention of preeclampsia and intrauterine growth restriction (PREDO) Study	Finland	Pseudonymized ID-coded data are stored at the Helsinki and Uusimaa Hospital district (HUS) and University of Helsinki virtual drives.	Longitudinal cohort study with collection of clinical data and biosamples. E.g data of psychological test results (questionnaires, interviews, individual tests); data from medical records and from the following Finnish Nationwide	Recruitment already completed; Biological Samples collected. Some of the biological data are generated.	Clinical data, biological samples and already generated data will be used in HappyMums implementation (e.g WP2,3,4).	According to the regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, and the Finnish legislation we have filled in the "Processor's record of processing activities" and the organization's (HUS) "Record of processing activities". In these records we have named persons who can access person-identified data, who can access the pseudonymized data, where data are being	According to Finnish legislation individual-level genotype data cannot be archived to public data repositories or shared as individual participant data. According to Finnish legislation data from nationwide registers can be shared with individual researcher with permission from the register authority and the data repository needs to be audited to conform requirements by findata.fi. Any individual-participant-data sharing needs to be approved also by HUS and UH. Any kind of individual participant data cannot be stored in open	HUS 482/E7/2004 (18.1.2005) HUS/1966/2021 (13.02.2021)

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
				Registers: Medical Birth Register (MBR), Care Register for Health Care (HILMO), Register on Reimbursement Medication (KELA) and Statistics Finland (SF); data from biological tissue samples and laboratory test results.			stored and how will secured access be controlled. Subject information (i.e. personal identifiers) is stored separately from the data; registry keeper is HUS. Researchers using data are delivered pseudonymized end-user data spreadsheets with no personal identifiers of the subjects (subject ID numbers are used as sole identifiers of the subjects). Only approved researchers will be able to gain access to the data, and they can only gain access to pseudonymized and only after signing a data confidentiality agreement.	repositories, even if fully anonymized. Mitigation measures: to share data as summary statistics, UH runs scripts provided for us and shares summary statistics, or UH requests required permissions and individual researcher conducts analyses on site.	
UH	The Intrauterine sampling in early pregnancy study (ITU)	Finland	Pseudonymized ID-coded data are stored at the Helsinki and Uusimaa Hospital district (HUS) and University of Helsinki virtual drives.	Longitudinal cohort study with collection of clinical data and biosamples. Data of psychological test results (questionnaires, interviews, individual	Recruitment already completed; Biological Samples collected. Some of the biological data are generated.	Clinical data, biological samples and already generated data will be used in HappyMums implementation (e.g WP2,3,4).	According to the regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, and the Finnish legislation we have filled in the "Processor's record of processing activities" and the organization's (HUS) "Record of processing	According to Finnish legislation individual-level genotype data cannot be archived to public data repositories or shared as individual participant data. According to Finnish legislation data from nationwide registers can be shared with individual researcher with permission from the register authority and the data repository needs to be	HUS 269/13/03/00/2009 (18.05.2010)

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
				tests); data from medical records and from the following Finnish Nationwide Registers: Medical Birth Register (MBR), Care Register for Health Care (HILMO), Register on Reimbursement Medication (KELA) and Statistics Finland (SF); data from biological tissue samples and laboratory test results, including DNA.			activities". In these records we have named persons who can access person-identified data, who can access the pseudonymized data, where data are being stored and how will secured access be controlled. Subject information (i.e. personal identifiers) is stored separately from the data; registry keeper is HUS. Researchers using data are delivered pseudonymized end-user data spreadsheets with no personal identifiers of the subjects (subject ID numbers are used as sole identifiers of the subjects). Only approved researchers will be able to gain access to the data, and they can only gain access to pseudonymized data and only after signing a data confidentiality agreement.	audited to conform requirements by findata.fi. Any individual-participant-data sharing needs to be approved also by HUS and UH. Any kind of individual participant data cannot be stored in open repositories, even if fully anonymized. Mitigation measures: to share data as summary statistics, UH runs scripts provided for us and shares summary statistics, or UH requests required permissions and individual researcher conducts analyses on site.	
EMC	Generation R and GenerationR Next	The Netherlands	The Netherlands (via Research Suite and	Longitudinal cohort study (incl. imaging data, clinical	Recruitment completed but data collection ongoing	Clinical data, biological samples and already generated data	Controlled open access system for data use and mandatory confidentiality agreements and Good	In case of a potential ethical issue, such as a data breach, Erasmus MC rules for security incidents are followed. This	All rounds of the Generation R Study have been approved by the medical ethics committee

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
			through EMC facilities)	data, and omics). Data is collected via interviews, questionnaires, and biological tissue samples. Data are also collected through linkage with external datasets, e.g.GP, pharmacy, hospital or school records (only done with consent from the participants).	(follow-up visits).	will be used in *HappyMums* implementation (e.g WP3 and WP4)	Clinical Practice training for project personnel. Datasets for researchers are pseudonymized. Data are shared with external parties only with management team approval, after signing a data transfer agreement and with project-specific pseudonymization.	includes informing the study coordinator, principal investigator, privacy contact person (dataleak team) and making an internal report.	of Erasmus MC, University Medical Center Rotterdam. Generation R also reports on a yearly basis to the medical ethics committee of Erasmus MC on the study progress. Generation R phase 1 (fetal) approval date: 17 December 2001, phase 2 (0-4 years) approval date: 9 January 2003, phase 3 (5 years) approval date: 28 February 2008, phase 4 (9 years) approval date: 17 October 2012, phase 5 (13 years) approval date: 9 February 2016, phase 6 (17 years) approval date: 30 April 2020, Generation R Next approval date: 24 September 2021.
AU	Brazilian High Risk Cohort Study (BHRCS)	Brazil	To be determined, in discussion now, as of ethical clearance in April 2023.	Longitudinal cohort study with collection of clinical data and biosamples	Recruitment already completed; Biological Samples collected. Some of the biological data are generated.	Clinical data, biological samples and already generated data will be used in *HappyMums* implementation (WP 4).	Protocol follows all requirements for data protection according with Brazilian Ethical Council http://conselho.saude.gov.br/images/comissoes/conep/documentos/NORMAS-RESOLUCOES/466_english.pdf	There may be risks concerning the tasks to be accomplished. Parents/guardians and their children may get tired of filling in the questionnaires. They may also feel anxious or embarrassed by answering questions about their own feelings and day-to-day behavior, because the contents involve emotions, and behaviors that can be	Ethical National Council of Brazil approved protocol 74563817.7.1001.5327 under report number 5.981.882 in 09/04/2023 authorizing clinical and biological samples transfer to AU and/or University of Milan to have epigenetic analysis done.

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
								unpleasant. We will try to minimize these possible effects by using trained evaluators. If the young person was selected for the MRI scan, there is a loud noise during the scan. To ease the discomfort, we provided earmuffs. In the event of any intercurrence or damage resulting from the research, the participant will receive all the necessary assistance funded by the project, at no personal cost to the participant or his/her responsible.	
HappyMums ongoing studies: Already existing cohorts, where the recruitment is still ongoing, and biological samples are being collected.									
UH	Implementation of prevention and treatment intervention of maternal depression during pregnancy (Imprint; Ilo Odottaa)	Finland	Pseudonymized ID-coded data are stored at the Helsinki and Uusimaa Hospital district (HUS) and University of Helsinki virtual drives.	Cluster-randomized clinical trial with collection of clinical data and biosamples. Data of psychological test results (questionnaires, interviews, individual tests); data from medical records and from the	Recruitment ongoing	Clinical data, data from medical records and nationwide registers, biological samples that will be generated will be used in HappyMums implementation (e.g WP7).	According to the regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, and the Finnish legislation we have filled in the "Processor's record of processing activities" and the origanization's (HUS) "Record of processing activities". In these records we have named persons who can access person-identified data, who can access the	According to Finnish legislation individual-level genotype data cannot be archived to public data repositories or shared as individual participant data. According to Finnish legislation data from nationwide registers can be shared with individual researcher with permission from the register authority and the data repository needs to be audited to conform requirements by findata.fi. Any individual-participant-data sharing needs to be approved also by HUS and UH. Any kind	HUS/841/2019 (14.10.2020, 25.5.2022) ClinicalTrials.gov (NCT04069091)

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
				following Finnish Nationwide Registers: Medical Birth Register (MBR), Care Register for Health Care (HILMO), Register on Reimbursement Medication (KELA) and Statistics Finland (SF); data from biological tissue samples and laboratory test results, including DNA.			pseudonymized data, where data are being stored and how will secured access be controlled. Subject information (i.e. personal identifiers) is stored separately from the data; registry keeper is HUS. Researchers using data are delivered pseudonymized end-user data spreadsheets with no personal identifiers of the subjects (subject ID numbers are used as sole identifiers of the subjects). Only approved researchers will be able to gain access to the data, and they can only gain access to pseudonymized ID-coded data without personal identifiers, and only after signing a data confidentiality agreement forms.	of individual participant data cannot be stored in open repositories, even if fully anonymized. Mitigation measures: to share data as summary statistics, UH runs scripts provided for us and shares summary statistics, or UH requests required permissions and individual researcher conducts analyses on site.	
UMIL	PRESeNT	Italy	Pseudonymized ID-coded data are stored at the project coordinator site IRCCS	Longitudinal cohort study with collection of clinical data and biosamples. Data of psychological	Recruitment and biological samples collection still ongoing.	Clinical data and biological samples will be used in HappyMums (WP2, WP3, WP4, WP7)	According to the regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. Subject information (i.e. personal identifiers) is	The study is working following all the relevant guidelines with regard to data collection design & safety, management and documentation. The institutional DPO at IRCCS Fatebenefratelli is part of the	

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
			Fatebenefratelli (Brescia). A project repository compliant with safety legislation is collecting all project data together	test results (questionnaires); data from biological samples and laboratory test results, including biochemical (immunological, hormonal), gene expression, omics.			stored separately from the data. Researchers using data are delivered pseudonymized end-user data spreadsheets with no personal identifiers of the subjects (subject ID numbers are used as sole identifiers of the subjects). Only approved researchers will be able to gain access to the data, and they can only gain access to pseudonymized ID-coded data without personal identifiers.	Ethics Committee and is constantly monitoring the study. Participants are being thoroughly informed of all necessary procedures and about all type of biological samples they are asked to provide. They are also being fully informed and asked consent for any type of information and personal data they need to provide. Specifically, they are also being asked to sign a Privacy Policy form. All these documentations have been reviewed and accepted by a local Ethical Committee.	
Clinical registers: Already existing databases and clinical resources used to investigate the association between available risk or protective factors and clinically relevant outcomes regarding depressive symptomatology during pregnancy.									
KCL	South London and Maudsley NHS Foundation Trust (SLaM)	UK	Data is currently stored on secure NHS server.	Clinical patient records	Data is stored on an ongoing basis, but an application would need to be made to conduct research on this dataset.	Clinical data to be used in H*appyMums* WP3.	Data fully anonymised and only accessible by those who have formally applied.	Data sharing will not be possible. KCL team can apply to hospital governance to access data and perform interrogations for a specific research question. They will then share summary statistics for purposes of HM.	Application would need to be made, once research questions and variables are defined.
UMIL	Population-based registry, known as	Lombardy, the most populated	The DWH uses a star schema, containing a set	Data from the social and health sector,	Data is stored on an ongoing basis, but an	Clinical data to be used in H*appyMums* WP3.	Data are fully anonymised immediately after their acquisition by the Region.	Data can be used for public health interest, planning, monitoring, services	A project proposal application would need to be made, once research

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
	Regional Data Warehouse (DWH).	Italian region, with 10 M inhabitants.	of large central repositories. These repositories contain the main core of data, without any redundancy.	anonymised immediately after their acquisition and restructured in a single repository.	application would need to be made to conduct research on this dataset.		Then, data are extracted, pre-processed, integrated and restructured in a single repository and only accessible by those who have formally applied.	evaluation, and operational research. For this purpose, after an additional process of anonymisation, Lombardy provides data for universities and other scientific institutes which elaborate studies and projects approved by the regional administration. Before the processing of personal data, Lombardy Region carries out a privacy impact assessment.	questions and variables are defined. Once the accreditation and the project proposal are accepted, the institution and Lombardy Region sign an agreement that states the aims of partnership, the duties of each sides involved, and the duration of the partnership based on the needs of the project. Then, the Lombardy Region sets up the virtual environment to access the DWH
OSR	OSR will analyse available cohorts among *HappyMums* consortium.	Italy	OSR will provide to partners pipelines and scripts to perform analyses locally. Only if required, data will be stored on OSR secure server, only available to those with approved access	Demographical, clinical and biological data in cohorts' studies as describe by partners.	Recruitments of cohorts are on charge of the research partners, as specified.	OSR will analyse data provided by partners for HappyMums WP2, WP3, WP4.	OSR will analyze fully anonymized data. Data will be only accessible by those who have formally applied.	When possible OSR will exploit the Federated Learning Platform, DataSHIELD or will provide tools to partners for federated analyses. Otherwise, they will ask for summary statics or deliver scripts among partners. Only if strictly necessary to achieve aims of WP2,3,4, OSR will directly ask partners data sharing. Data will be pseudoanonimized and codes will be kept by pertinent clinical partners, transferred and stored according to EU guidelines and defined protocols among the consortium.	Using secondary data generated by partners', OSR will rely on their ethical reviews, treating data accordingly. Data Transfer Agreement will be defined among OSR and partners.

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
HappyMums Clinical Study: Data collected in the feasibility study that will be run in *HappyMums.*									
ABACUS	*HappyMums* Clinical Study data from digital tools	Italy, Croatia, German, Finland, UK, Poland	ABACUS cloud infrastructure	Digital data generated during clinical study using *HappyMums* mobile app.	To do (2024)	Digital data generated during clinical study using *HappyMums* mobile App	Protected network, authentication, regular backups, encryption, data masking, employee education, data breach procedures	Presence of personal data: → MITIGATION: data encryption, data masking and authentication. Incidental findings: → MITIGATION: sharing of data with clinical partners.	Ethical approval to be requested by clinical partners. Types of data: relational database with tabular data, time series, text files, audio files.
KCL, UH, UMIL, OSR, CHARITE, CUC, SWPS	HappyMums Clinical Study	Multiple	Mobile app data storage will be managed by the ABACUS (Milan, Italy). Data from questionnaires, interviews, etc. will be stored within the RedCap system on the server at stored at individual centres (no material transfer planned), or by ABACUS when submitted via app	Cohort study with collection of clinical data and biosamples	Protocol design in progress – led by KCL but with input from all other centres. Clinical study to begin March 2024.	Testing the mobile application. Contributing to the *HappyMums* Task 5.4.	Double tables will be organized – one with confidential data and the other with anonymized data. Only the main investigators of each center will have the access to the confidential tables. Only the anonymized data will be shared outside each partner. Redcap will be used for data recruitment and collection.	Disclosing personal and sensitive data. Identifying severely depressed and suicidal women – these women will be contacted with additional information where to find professional support and help.	Ethical approval will be obtained by each center once the *HappyMums* Clinical Study protocol is ready. We will seek consent either online or face-to-face, as to make participation in the study as hassle free as possible.
External resources used to collect more requirements and data regarding the digital platform developed in *HappyMums.*									
UB	AffectNet	N/A	Locally (UB)	Public (Image)	N/A	Pretrain emotion recognition	Third-party license compliance. http://mohammadmahoo	Protected attributes related biases: Data level: Data Balancing	N/A

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
						models from still images. WP8	r.com/affectnet-request-form/	Model level: Adversarial training.	
UB	D-vlog: Multimodal Vlog Dataset for Depression Detection	N/A	Locally (UB)	Public (Video)	N/A	Pretrain depression recognition models from video. WP8	Third-party license compliance.	Protected attributes related biases: Data level: Data Balancing Model level: Adversarial training.	N/A
UB	Audiovisual Emotion Challenge (AVEC)	N/A	Locally (UB)	Public (Video)	N/A	Pretrain depression recognition models from video. WP8	Third-party license compliance.	Protected attributes related biases: Data level: Data Balancing Model level: Adversarial training.	N/A
UB	Pittsburgh Depression Dataset	N/A	Locally (UB)	Public (Video)	N/A	Pretrain depression recognition models from video. WP8	Third-party license compliance.	Protected attributes related biases: Data level: Data Balancing Model level: Adversarial training.	N/A
UB	Crowd-sourced Emotional Multimodal Actors Dataset(CREMA-D)	N/A	Locally (UB)	Public (Video)	N/A	Pretrain emotion recognition models from video. WP8	Third-party license compliance. https://opendatacommons.org/licenses/odbl/1-0/	Protected attributes related biases: Data level: Data Balancing Model level: Adversarial training.	N/A
UB	Sewa Database	N/A	Locally (UB)	Public (Video)	N/A	Pretrain emotion recognition models from video. WP8	Third-party license compliance.	Protected attributes related biases: Data level: Data Balancing Model level: Adversarial training.	N/A

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
UB	AFWL	N/A	Locally (UB)	Public (video)	N/A	Pretrain landmark detection models from still images. WP8	Third-party license compliance. https://www.tugraz.at/index.php?id=17757	Protected attributes related biases: Data level: Data Balancing Model level: Adversarial training.	N/A
HappyMums Preclinical Studies: Data collected in the animal experiments carried out in *HappyMums*									
UMIL	Behaviour PNS/ELS model	Italy	Virtual drives	Behavioral data; data format: mp4, Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	According to Italian legislation, the request for ethical approvals of the animal license will be requested to the Animal Care committee of UMIL and to the Italian Ministry of Health. The license will be approved by 2023. The procedures will be performed according to the Italian ethical regulations of animal handling.
UMIL	RNA-seq PNS/ELS model	Italy	Virtual drives	Transcriptomic data; data format: fastQ, BAM, excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	According to Italian legislation, the request for ethical approvals of the animal license will be requested to the Animal Care committee of UMIL and to the Italian Ministry of Health. The license will be approved by 2023. The procedures will be performed according to the Italian ethical

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
									regulations of animal handling.
UMIL	Brain imaging PNS/ELS model	Italy	Virtual drives	Cross-sectional ex vivo brain imaging data; data format: Tiff, Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	According to Italian legislation, the request for ethical approvals of the animal license will be requested to the Animal Care committee of UMIL and to the Italian Ministry of Health. The license will be approved by 2023. The procedures will be performed according to the Italian ethical regulations of animal handling.
UMIL	Biochemical PNS/ELS model	Italy	Virtual drives	Biochemical data; data format: Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	According to Italian legislation, the request for ethical approvals of the animal license will be requested to the Animal Care committee of UMIL and to the Italian Ministry of Health The license will be approved by 2023. The procedures will be performed according to the Italian ethical regulations of animal handling.

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
UMIL	qPCR PNS/ELS model	Italy	Virtual drives	Gene expression data; data format: Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	According to Italian legislation, the request for ethical approvals of the animal license will be requested to the Animal Care committee of UMIL and to the Italian Ministry of Health. The license will be approved by 2023. The procedures will be performed according to the Italian ethical regulations of animal handling.
UZH	Behaviour MSUS and CSIR model	Switzerland	Virtual drives	Behavioral data; data format: mp4, Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	Experiments will follow Swiss regulations for animal experimentation and will be conducted under the animal licenses ZH021/2022 and ZH052/2022.
UZH	RNA-seq MSUS and CSIR model	Switzerland	Virtual drives	Transcriptomic data; data format: fastQ, BAM, excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	Experiments will follow Swiss regulations for animal experimentation and will be conducted under the animal licenses ZH021/2022 and ZH052/2022.
UZH	qPCR and DNAme MSUS and CSIR model	Switzerland	Virtual drives	Gene expression and DNA methylation data; data	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	Experiments will follow Swiss regulations for animal experimentation and will be conducted under the animal licenses

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
				format: Excel sheets					ZH021/2022 and ZH052/2022.
UZH	Brain imaging MSUS and CSIR model	Switzerland	Virtual drives	Cross-sectional ex vivo brain imaging data; data format: Tiff, Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	Experiments will follow Swiss regulations for animal experimentation and will be conducted under the animal licenses ZH021/2022 and ZH052/2022.
UZH	Biochemical MSUS and CSIR model	Switzerland	Virtual drives	Biochemical data; data format: Excel sheets	N/A	All data will be used to achieve tasks of WP6 and WP7	Third-party license compliance. Data on secure server with restricted access	N/A	Experiments will follow Swiss regulations for animal experimentation and will be conducted under the animal licenses ZH021/2022 and ZH052/2022.
UvA	Biochemical and morphological PNS/ELS model	The Netherlands	Virtual drives	Biochemical and morphological data; data format: Tiff, Excel sheets	N/A	All data will be used to achieve tasks of WP6	Third-party license compliance. Data on secure server with restricted access	N/A	All experimental studies are approved by the Animal Ethics Committee of University of Amsterdam. The license will be requested by 2023.
COLUMBIA	Cognitive flexibility training in mice	USA	External hard drive in the laboratory	Behavioral Data format: mp4, Excel sheets	N/A	All data will be used to achieve tasks of WP7	Laboratory computers are part of a 10/100/1000Mbps TCP/IP based Cisco network that resides behind a Netscreen firewall	N/A	Institutional Animal Care protocol AC-AABW3652 (April 6th, 2023) includes prenatal and early life stress procedures and cognitive flexibility training in mice.

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
KCL	PNS+LBN, MSUS, and SIR model ex vivo MR imaging	UK	KCL Neuroimaging network servers (NaN)	Cross-sectional ex vivo brain imaging data; data format: Tiff, Excel sheets	Generation of animal models is underway at partner sites. MR Imaging will commence in June 2023 and run until the end of the grant for WP6 and 7	Animal model data for WP6 and WP7	Data is on a secure network with restricted access and backup systems offsite.	N/A	Animal experiments are conducted with approval from the local King's College London ethics committee in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986.
WUR	RNA-Seq and DNA me FCSIR and FPNS fish model	Netherlands	Virtual drives	Transcriptomic and DNA methylation data; data format fastQ, BAM, excel sheets	Breeding experiment in progress	All data will be used to achieve tasks of WP6	Data produced will be stored according to Wageningen University (WU) Regulations, following accepted standards of scientific conduct, as laid down in the VSNU report on the Netherlands Code of Conduct on Scientific Practice (2005). These regulations safeguard transparency in data production and management. During the project, data sets are safely stored on protected servers at WU. WU & Research has set up several tailored services via the Data-management support hub.	N/A	All experimental studies are approved by the Animal Ethics Committee of Wageningen University and Research (2020.W-0027).

Partner	Dataset name	Country	Where data will be stored	Nature of data	Recruitment status	Relation to *HappyMums*	Description of security measures	Potential ethical issues and proposed mitigation measures	Ethical reviews or agreements that have been fulfilled, including dates
WUR	Anatomical, histological, biochemical FCSIR and FPNS fish model	Netherlands	Virtual drives	Anatomical, histological, biochemical data; data format: Tiff, Excel sheet	Breeding experiment in progress	All data will be used to achieve tasks of WP6	Data produced will be stored according to Wageningen University (WU) Regulations, following accepted standards of scientific conduct, as laid down in the VSNU report on the Netherlands Code of Conduct on Scientific Practice (2005). These regulations safeguard transparency in data production and management. During the project, data sets are safely stored on protected servers at WU. WU & Research has set up several tailored services via the Data-management support hub.	N/A	All experimental studies are approved by the Animal Ethics Committee of Wageningen University and Research (2020.W-0027).