D6.5 – Study 1 initiation package

Revision: v.3.5.0

Work package	WP6
Task	Task 6.2
Due date	30.04.2025
Submission date	30.04.2025
Deliverable lead	UKCM (university Medical Center Maribor), UM (University of Maribor as clinical coordinator)
Version	3.5
Authors and Contributors	Izidor Mlakar (UM), Urška Smrke (UM), Ana Dimič (UM), Evangelia Charmandari (NKUA), Penio Kassari (NKUA), Martin Bigec (UKCM), Jernej Dolinšek, (UKCM) Jernej Vidmar (UKCM), Sonja Golob Jančič (UKCM), Mojca Podgoršek (UKCM), Kaja Golija (UKCM), Tjaša Hertiš Petek (UKCM), Eirini Bathrellou (HUA), Billy Langlet (KI), Athena Balaska (PENTELI), Fafoula Olga (PENTELI), Radka Savova (BLOCKS), Elitsa Brankova (BLOCKS), Svetlin Hansov (BLOCKS), (CHUL), Caroline Gernay (CHUL), Billy Langlet (KI), Alkyoni Glympi (KI), Ioannis Ioakimidis (KI), Vasileious Papapanagiotou (KI), WLC?, CHUL?, VHIR?
Reviewers	Giada Martello (Martel); Danai Kyrkou (ViLabs)

Co-funded by the European Union

 © 2023-2025 BIO-STREAMS

Abstract	The deliverable Study 1 Initiation Package, provides the operational and methodological framework for the retrospective study titled “Using electronic health records and AI to improve health outcomes for children with obesity across Europe” (ISRCTN12357025).
Keywords	Obesity prevention, Artificial intelligence, Machine learning, Virtual biobank, Retrospective study, Study initiation package

DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
1.0	17.01.2025	Study 1 initiation package - outline	Izidor Mlakar
2.0	14.02.2025	Initial Content Added	Izidor Mlakar, Martin Bigec
2.5	10.03.2025	Partner Contributions	All Clinical Partners
2.8	26.03.2025	Contributions from WLC, UOI	Ana Rehberger, Peter Davis, Eleni Georga
3.0	07.04.2025	Final contributions, deliverable sent of internal review	Izidor Mlakar, Martin Bigec All partners
3.5	22.04.2025	Final refinement, deliverable sent for revision by VILABS and MARTEL , Deliverable ready for final quality check by coordinator	Izidor Mlakar, Giada Martello, Danai Kyrkou

Disclaimer

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the other granting authorities. Neither the European Union nor the granting authority can be held responsible for them.

Copyright notice

© 2023 - 2025 BIO-STREAMS Consortium

Project co-funded by the European Commission in the Horizon Europe Programme	
Nature of the deliverable:	D6.5 Study 1 initiation package
Dissemination Level	

PU	*Public, fully open, e.g. web*	**X**
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

EXECUTIVE SUMMARY

The BIO-STREAMS project is a multidisciplinary, four-year initiative funded by the European Union, aimed at tackling the urgent public health challenge of childhood and adolescent obesity across Europe. Recognizing the complex interplay of genetic, environmental, and lifestyle factors underlying obesity, the project seeks to develop a comprehensive digital ecosystem to support prevention, management, and research in this field.

This deliverable, Study 1 Initiation Package, provides the operational and methodological framework for the retrospective study titled “Using electronic health records and AI to improve health outcomes for children with obesity across Europe” (ISRCTN12357025). The study’s primary objectives are to identify new sub-cohorts and biological pathways related to obesity and to validate the predictive power of artificial intelligence (AI) algorithms in assessing obesity risks. The research will analyze harmonized electronic health records from seven clinical sites in six EU countries, encompassing data from over 8,800 children aged 5–18 years.

The study is structured in sequential phases, from protocol development and data extraction to model validation and dissemination of results, with clearly defined timelines and milestones. The project emphasizes continuous ethical monitoring, risk management, and user-centered design to maximize real-world applicability and impact.

 © 2023-2025 BIO-STREAMS

Table of contents

 © 2023-2025 BIO-STREAMS

 © 2023-2025 BIO-STREAMS

 © 2023-2025 BIO-STREAMS

List of figures

 © 2023-2025 BIO-STREAMS

List of tables

 © 2023-2025 BIO-STREAMS

Abbreviations

AI	**Artificial Intelligence**
EQF	**European Qualifications Framework**
EU	**European Union**
FAIR	**Findable, Accessible, Interoperable, and Reusable (data principles)**
GA	**Grant Agreement**
GDPR	**General Data Protection Regulation**
GWAS	**Genome-Wide Association Study**
HCI	**Human-Computer Interaction**
HPC	**High Performance Computing**
ICC	**Intraclass Correlation Coefficient**
ICD10	**International Classification of Diseases, 10th Revision**
ISRCTN	**International Standard Randomised Controlled Trial Number**
IP	**Internet Protocol**
IT	**Information Technology**
KPI	**Key Performance Indicator**
LOINC	**Logical Observation Identifiers Names and Codes**
LIME	**Local Interpretable Model-Agnostic Explanations**
M	**Month**
ML	**Machine Learning**
NGO	**Non-governmental Organization**
NCDs	**Non-Communicable Diseases**
OMOP	**Observational Medical Outcomes Partnership**
SME	**Small and Medium-sized Enterprise**
TCP	**Transmission Control Protocol**
WP	**Work Package**

 © 2023-2025 BIO-STREAMS

1 Introduction

BIO-STREAMS, here referred to as the '**Project**,' is a multi-disciplinary EU project, primarily focusing on addressing childhood and adolescent obesity. The Project's objectives, components, and deliverables are manifold. To begin with, it aims to develop and implement three key elements:

- EU-wide Childhood/Adolescence Obesity Biobank, designed to facilitate data sharing across the EU for research and innovation purposes
- Accessible Obesity Platform, utilizing the resources of the Biobank to offer various services, including applications
- EU Community Network dedicated to Childhood/Adolescence Obesity

The Project spans across four years, from May 2023 (M1) to April 2027 (M48).

Childhood obesity is a pressing public health issue that poses serious health risks to millions of children and adolescents across Europe and globally. The complex etiology of obesity, involving interplay between genetic, environmental, and lifestyle factors, requires multifaceted interventions. In response to this challenge, the BIO-STREAMS project was initiated to develop a digital platform designed to support the prevention and management of obesity among young populations. This platform aims to integrate the latest research and evidence-based practices into tools that are both practical and accessible for healthcare providers, educators, policymakers, and families.

This document describes and defines the Study Initiation Package for the retrospective study "Using electronic health records and AI to improve health outcomes for children with obesity across Europe" (ISRCTN12357025), which provides a comprehensive framework for launching and managing the research. It serves as a central resource for all pilot sites, encompassing essential information, such as the study's objectives, design, and methodology, as well as strategies for participant recruitment, ensuring regulatory compliance, and outlining operational management.

1.1 Scope of Project

The scope of the Project's is detailed within the GA and Horizon Europe Proposal ('**Proposal**'). [1] This section offers a concise overview, emphasizing the Project's most pertinent aspects for this Report.

As mentioned earlier, BIO-STREAMS is a diverse project aimed at addressing childhood and adolescent obesity through the utilization of data and technology. The Proposal includes a visual representation of the BIO-STREAMS 'Ecosystem,' summarizing its various elements.[2]

---

[1] *Proposal for Horizon Europe Work Programme 2021 – 2022 / Health HORIZON-HLTH-2022-STAYHLTH-01-05-two-stage - Prevention of obesity throughout the life course Multi-Pillar Framework for children Anti-Obesity Behaviour building on an EU Biobank, Micro-Moments and Mobile Recommendation Systems. Ref. Ares (2022) 6184745 - 07/09/2022.*

[2] *Proposal, 7.*

© 2023-2025 BIO-STREAMS

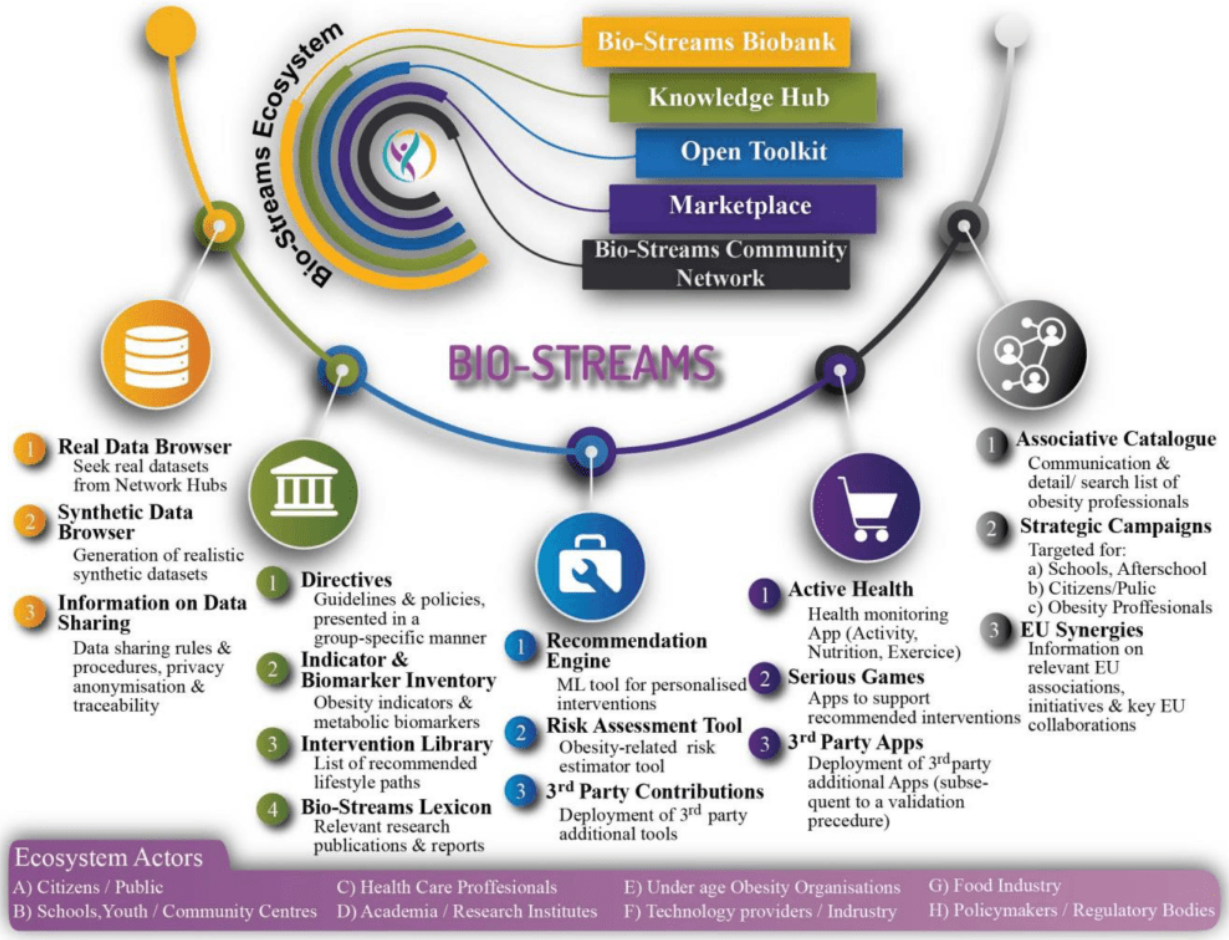

Figure 1: A visual representation of BIO-STREAMS Ecosystem

The BIO-STREAMS 'Ecosystem' incorporates five main components, supported by the necessary backbone services & packages',[3] which are presented in the table below:[4]

[3] *Proposal, 6.*

[4] *Proposal, 7.*

© 2023-2025 BIO-STREAMS

Table 1: BIO-STREAMS backbone services and packages

Service/Package	Short Description
Bio-Streams Node Bundle (BNB)	It includes hardware & software resources, installed at every *Bio-Streams* member site to create a Data Hub. All Hubs formulate a Hub Network constituting the *Bio-Streams* Biobank.
Synthetic Data Generator (SDG)	Service for on-demand creation of synthetic datasets based on knowledge extracted from real data in the *Bio-Streams* Biobank.
Data Handler	Service for dataset cataloguing, browsing, requesting, sharing & tracking. Operation will be governed by the *Bio-Streams* regulatory framework and Data Management Plan.
Security Assurance Framework (SAF)	SAF will ensure secure and privacy-preserving operation of all *Bio-Streams* components and services, dynamically operating both synchronously and asynchronously.
ML tools	Federated ML supporting knowledge representation & extraction (e.g. SDG), risk assessment and pathway recommendation.

Structurally, the Project is organized into five 'Phases', which are summarized from the Proposal[5] as follows:

1. Build knowledge foundation for the design, development and implementation of BIO-STREAMS
2. Build BIO-STREAMS Biobank
3. Integrate BIO-STREAMS Solution
4. Create BIO-STREAMS Community Network
5. Validate and Evaluate the BIO-STREAMS Solution in multiple Pilot-Sites

While each Phase holds significance for the Project's overarching goals, the Biobank stands as a cornerstone, essential for advancing other Phases or components within the BIO-STREAMS Ecosystem. The Project entails gathering and examination of biological samples. However, it is the digitized data that will primarily drive the Project toward its objectives, encompassing metadata extracted from biological material.[6] Therefore, the BIO-STREAMS Biobank is best conceptualized as a virtual biobank, serving as an 'EU-wide data-sharing centre'[7] for childhood/adolescent obesity, housing various data types such as demographic, behavioral, clinical, genetic/epigenetic, and cost data.[8] The critical role of the BIO-STREAMS Biobank within the Project is depicted in the accompanying graphic:[9]

[5] *Proposal, 7-18.*

[6] *Proposal, 12. Note that the GDPR distinguishes biological material from data that is derived from it (at recital 33).*

[7] *Proposal, 2.*

[8] *Ibid.*

[9] *Proposal, 12.*

© 2023-2025 BIO-STREAMS

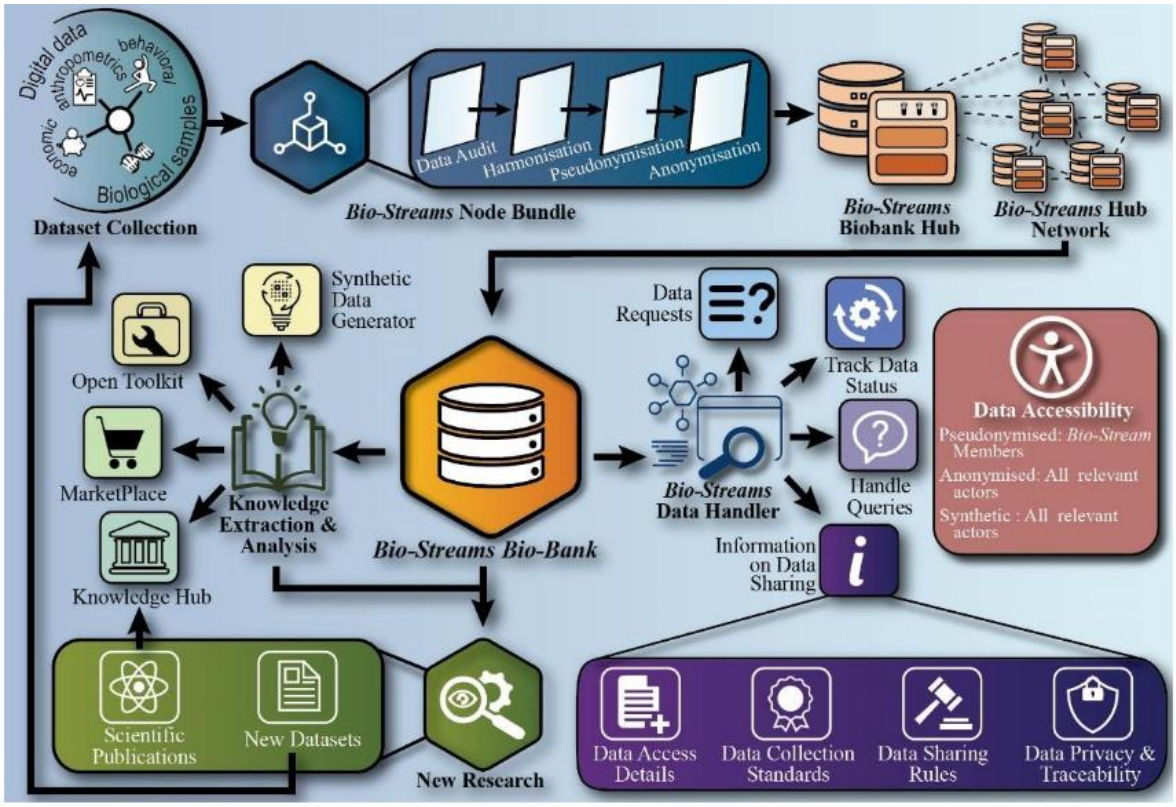

Figure 2: Process of collecting, storing, analyzing, and sharing Project data

The graphic above illustrates the process of collecting, storing, analyzing, and sharing Project data in connection with the BIO-STREAMS Biobank.

1.2 Scope of Report

The Study Initiation Package (SIP) is designed to provide a complete overview and guide for initiating and conducting the Retrospective Study, titled “Using electronic health records and AI to improve health outcomes for children with obesity across Europe” (https://doi.org/10.1186/ISRCTN12357025). It defines critical aspects necessary for the successful launch and management of the clinical study. The document serves as a central reference point for investigators, sponsors, regulatory bodies, and other stakeholders involved in the trial execution.

Key areas that this document includes are:

1. Study Overview: Provides a concise summary of the trial, including its objectives, design, and key endpoints
2. Scientific Background: Contextualizes the study within current medical and scientific knowledge
3. Detailed Study Design: Outlines the investigational plan, methodology, and specific procedures to be followed
4. Participant Recruitment: Presents a comprehensive strategy for identifying, engaging, and enrolling suitable study participants
5. Regulatory Compliance: Addresses ethical considerations and ensures alignment with Good Clinical Practice (GCP) guidelines and relevant regulatory requirements

© 2023-2025 BIO-STREAMS

6. Operational Framework: Delineates responsibilities, oversight mechanisms, and risk management strategies
7. Study Management: Includes timelines, milestones, and key performance indicators for monitoring study progress
8. Essential Documentation: Incorporates critical documents such as the full study protocol and other relevant documents and forms

© 2023-2025 BIO-STREAMS

2 Study Overview

Title	EU-Wide Childhood Obesity Biobank and Decision Support System: An Observational Cohort Study
Aim	The purpose of this study is to generate a homogenized dataset consisting of retrospective data on demographic, medical (healthcare indices), behavioral and family history information derived from the Electronic Health Records of the clinical sites to (1) identify biological pathways conferring childhood obesity and (2) to validate the predictive/prognostic power of Artificial Intelligence (AI) algorithms developed by BIO-STREAMS consortium.
Registration of the Study	ISRCTN12357025
DOI:	https://doi.org/10.1186/ISRCTN12357025
Retrospective study design (RSD)	Structured and pseudonymized retrospective data collected from 7 clinical sites. Blind tests to validate the AI algorithms to identify susceptibility and possible health outcomes.
RSD Primary objective	To create a holistic EU Obesity Virtual Biobank by combining retrospective data from multiple EU cohorts and allow for further analyses regarding risk factors and biomarkers of overweight in a large sample. To determine the best features as input to the AI models for predicting susceptibility and possible health outcomes.
RSD Secondary objective	The secondary aim of the study is to assess clinician feasibility of integrating the EU Childhood Obesity Platform into regular clinical routine.
RSD Primary endpoint	(1) Number of new obesity pathways discovered (2) Sensitivity and specificity of the AI models for prediction of risks for obesity (compared to the detailed clinical assessment).
RSD Secondary endpoint	(1) Suitability of the EU Childhood Obesity Platform measured through usability, acceptance, user experience, trust (related to AI predictions) (2) Homogenized retrospective BIO-STREAMS dataset (3) Data curation, data management solution, and infrastructure (4) Cost and economic benefit
Duration of clinical investigation	36 months
Subject population	In total, more than 8800 subjects (5600 children with overweight/obesity 3200 children with normal weight) are estimated to be included across all the study's 7 clinical sites: 1. University Medical Centre Maribor (UKCM)

© 2023-2025 BIO-STREAMS

2. National and Kapodistrian University of Athens (NKUA)
3. Karolinska Institute (KI)
4. Blocks Health and Social Care EOOD (BLOCKS)
5. Hospital Universitari Vall d'Hebron (VHIR)
6. Centre Hospitalier Universitaire de Liège (CHUL)
7. Penteli General Childrens' Hospital (PENTELI)

Sex distribution: In 2020, around 175 million children and adolescents aged 5-19 years worldwide were considered obese[10]. In the study female and male subjects will be recruited in proportions. However, it must be noted that boys are more susceptible to obesity than girls. A significant difference between genders in European cohorts was in favor of girls, i.e. fewer girls were obese than boys. On average across 23 EU countries, 14% of boys and 10% of girls aged 7-8 years were obese, according to the COSI study[11]. Overall, the prevalence of overweight and obesity among boys is 31%, while among girls it is 28%[12].

The data points of the following subject groups will be included in this retrospective study:

- Children and adolescents, age of participants: 5-18 years
- Children and adolescents with normal weight: BMI less than 1 SD above the median of the WHO growth reference for children and adolescent[13], not diagnosed with severe mental disorder, epilepsy, severe chronic medical conditions, cancer, orthopedic problems, known family predisposition, eating disorders or do not have any serious chronic disease.
- Children and adolescents with increased body weight (overweight) or with obesity: overweight defined as BMI more than 1 SD to 2 SD above the median of the WHO growth reference for children and adolescent; obesity defined as BMI more than 2 SD above the median of the WHO growth reference for children and adolescent not diagnosed with an eating disorder, severe mental disorder, epilepsy or severe chronic medical conditions, cancer prior to overweight/obesity.

To evaluate primary outcome, sensitivity and specificity data from 5% randomly selected retrospective subjects will be reserved for the validation of the models using blind tests. 30 healthcare experts and 30 researchers will participate. Experts need to be healthcare professionals involved in clinical research or practice obesity in the pediatric population. Researchers must be

[10] https://www.statista.com/statistics/1386146/number-of-obese-children-and-adolescents-worldwide-forecasts-by-gender/

[11] *https://www.oecd-ilibrary.org/sites/health_glance_eur-2018-26-en/index.html?itemId=/content/component/health_glance_eur-2018-26-en*

[12] *https://www.who.int/europe/news/item/03-03-2023-childhood-obesity--five-facts-about-the-who-european-region*

[13] WHO growth reference for children and adolescent: https://www.who.int/tools/growth-reference-data-for-5to19-years

© 2023-2025 BIO-STREAMS

	engaged in clinical research on Big Data and/or obesity (e.g. epidemiology of obesity, obesity research, etc.)
Number of subjects	In total, more than 8800 subjects are necessary to be included from 7 clinical sites to account for the minimum sample size per group (see sample size justification) and the relatively low incidence rate of obesity (< 6%) (UKCM, NKUA, KI, BLOCKS, VHIR, CHUL, PENTELI).
Number of Sites	7 clinical sites: 1) UKCM 2) NKUA 3) KI 4) BLOCKS 5) VHIR 6) CHUL 7) PENTELI
Sponsor information	University of Maribor (UM), Research organization University Medical Center Maribor (UKCM), Hospital/treatment Centre
External organizations involved in the clinical investigation	No organizations external to the BIO-STREAMS consortium will be involved in the clinical investigation. Each clinical site, contributing with data, will allocate a pilot leader (e.g. the Principal Investigators) to oversee the operations within the pilot study. Project's Clinical Manager, Izidor Mlakar (University of Maribor) will oversee the execution of all the studies. Oversight of the Data Operations will be carried out by Magdalena Góralczyk, Peter Davis and Lucrezia Nicosia (White Label Consultancy APS) and the Project's Data Protection Officer, Dimitris Kalogeras (ICCS). Each clinical partner also has their own DPO (contact details of each provided below).
Inclusion criteria	▪ Age of participants between 5 and 18 years ▪ (Overweight Cohort) BMI more than 1 SD to 2 SD above the median of the WHO growth reference for children adolescents ▪ (Obese Cohort): BMI more than 2 SD above the median of the WHO growth reference for children and adolescents ▪ (Normal Weight Cohort) BMI up to 1 SD above the median of the WHO growth reference for children and adolescents[14]
Exclusion criteria	▪ Age under 5 years or above 18 years ▪ BMI lower than 2 SD below the median of the WHO growth reference for children and adolescents ▪ diagnosed with an eating disorder, severe mental disorder, epilepsy or severe chronic medical conditions, cancer prior to overweight/obesity, untreated hypothyroidism

[14] WHO growth reference for children and adolescent: https://www.who.int/tools/growth-reference-data-for-5to19-years

© 2023-2025 BIO-STREAMS

Data Oversight	The study will process sensitive personal data of children, requiring rigorous data oversight and strict adherence to ethical and legal standards, particularly the General Data Protection Regulation (GDPR). To protect participant privacy, the study will implement advanced organizational and technical safeguards, such as encryption, ensuring that only necessary information is accessible for lawful research purposes. All data processing activities will be confined to the European Union. The personal data used in the study are part of the participants' existing medical records and will be stored accordingly. The legal basis for data processing relies on the research exemption under Article 89 of the GDPR, along with either the purpose limitation exemption (Article 5(1)(b)) or the public interest in scientific research (Articles 6(1)(e) and 9(1)(j)). Obtaining individual consent is not feasible due to the large number of subjects and associated logistical challenges. Throughout the study, appropriate safeguards—such as pseudonymization and secure data transfer—will be maintained, and these measures will be reviewed during the ethics approval process at each site. Structured and pseudonymized data will be transferred using specialized tools, with Data Transfer Agreements (DTA) and Data Processing Agreements (DPA) signed between all parties. All study-related documents will be retained by the consortium until the end of the BIO-STREAMS project.
Clinical investigation financing	This Clinical study is part of the BIO-STREAMS European Project funded within the research and innovation program of the Horizon Europe under N° 101080718. The funding source had no impact on the decision to carry out the study or on its design. The funding source will have no impact on the decisions related to publishing this research or its outcomes. The content of this document does not reflect the official opinion of the European Union or any other institution.
Person paying compensation	No compensation is provided for costs and time incurred in participating in a clinical investigation. The clinical sites will consider compensation for travel costs.

© 2023-2025 BIO-STREAMS

3 Study Information

The details of the study are defined in Appendix A: Detailed Study Protocol. In this section we provide an overview of the study.

3.1 Background

The rising prevalence of childhood obesity has become a significant public health concern worldwide, associated with increased risks of metabolic disorders, psychological issues, and long-term health complications. Social determinants such as globalization and sedentary lifestyles contribute to obesity by promoting unhealthy behaviors. Despite advances in genetics, much of the variability in body mass index (BMI) remains unexplained, highlighting the need for comprehensive approaches to understanding and addressing obesity.

With increasing volumes of electronic health data becoming available, researchers are applying big data approaches to track obesity prevalence and identify risk factors[1]. Machine learning and AI offer promising pathways for screening, diagnosis, and personalized intervention development, potentially addressing the heterogeneity in obesity's pathophysiology by discovering complex patterns in high-dimensional data.

3.2 Study Objectives

The primary objectives of the BIO-STREAMS project, are to discover new sub-cohorts and biological pathways related to obesity and validate the predictive power of AI algorithms in identifying obesity risks.

The study is driven by the following primary research questions:

(1) What are the clinical, biochemical, and lifestyle/behavioural predictors of obesity?

(2) What is the minimal set of variables required to train AI models that can accurately predict clinical pathways in childhood obesity?

(3) Can the different interactions create new cohorts underlining more efficient and patient-centered interventions?

(4) How do clinicians evaluate the feasibility of using the system in clinical routine, based on usability, trust and acceptance, and user experience while using the EU Childhood Obesity Platform?

Primary outcomes include:

(1) Number of new obesity pathways discovered

(2) Sensitivity and specificity of the Machine/Deep Learning models for prediction of risks for obesity

Secondary outcomes include:

(1) Suitability of the EU Childhood Obesity Platform measured through usability, acceptance, user experience and trust (related to AI predictions)

(2) Data curation, data management solution, and infrastructure

(3) Homogenized retrospective BIO-STREAMS dataset

(4) Cost and economic benefit

© 2023-2025 BIO-STREAMS

3.3 Intervention

No direct interventions are performed; the study focuses on analysing existing data to generate insights and predictive models for childhood obesity. The study includes the following methodological phases:

(1) Data Collection: Retrospective data extraction from Electronic Health Records of 7 European clinical sites, including demographic, clinical, biochemical, and lifestyle/behavioral variables for children aged 5-18 years.

(2) Data Harmonization: Transformation of collected data into a unified format using a common data model and ontology to create a standardized EU-wide childhood obesity virtual biobank.

(3) Data Analysis: Application of statistical analyses and machine learning techniques to identify risk factors associated with childhood obesity and its health-related outcomes.

(4) AI Model Development: Creation of predictive/prognostic AI systems using the harmonized dataset to identify obesity risk factors and project clinical pathways.

(5) Model Validation: Randomized blind tests using 5-10% of the data to validate the sensitivity and specificity of the AI models in predicting obesity risks and outcomes. Blind tests are conducted to compare AI algorithm results with detailed clinical assessments. 30 clinicians and 30 researchers will participate.

(4) Platform Evaluation: Assessment of the EU Childhood Obesity Platform's feasibility, usability, and economic impact through evaluations by clinicians and researchers.

(6) Cohort Comparison: Analysis of data from both overweight/obese (n = 5600) and normal weight (n = 3200) children to identify distinguishing factors and potential intervention points.

3.4 Study and Execution Plan

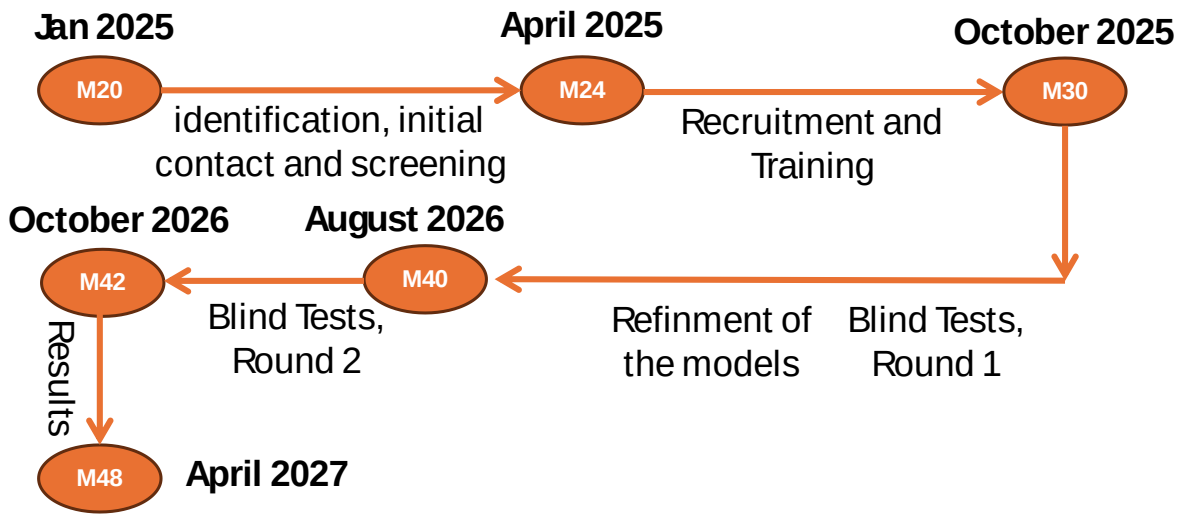

*Figure 3: Retrospective Study Timeline, *M refers to the month of the project and not the study.*

.

© 2023-2025 BIO-STREAMS

Figure 3 outlines the initial study execution plan with the following key activities (M refers to the month of the project and not the study):

1. **January 2025 (M20) to April 2025 (M24):**
   - Initial contact and screening activities for internal and external experts (clinicians and researchers).
   - All data operations are completed during this period, including identification of data, mapping to common semantic model, and harmonization.
   - All relevant DTA/DPAs are signed and come in force.

2. **April 2025 (M24) to October 2025 (M30):**
   - Recruitment and training of the 30 clinicians and 30 researchers. By M27 (July 2025) 50% of participants need to be recruited.
   - At the same time the final data export, data analysis and AI model refinement with the retrospective data will be carried out.

3. **October 2025 (M30) to August 2026 (M40):**
   - First round of blind tests to validate the AI models, this includes: (1) Test Data Preparation (5-10% of dataset) - First 3 weeks of October, (2) Evaluation by 30 Clinicians & 30 Researchers - Late October through November, (3) Comparison with Clinical Assessments - Mid-November through December.
   - Refinement of the models - the iteration and improvement of the AI models based on the first round of testing.
4. **August 2026 (M40) to October 2026 (M44):**
   - Second round of blind tests to validate the final version of models, this includes (1) Refined Model Test Data Preparation - First 3 weeks of August, Evaluation by 30 Clinicians & 30 Researchers - Late August through September and (3) Sensitivity & Specificity Assessment - Late-September through October.

5. **October 2026 (M44) to April 2027 (M48):**
   - Final analysis and compilation of results.

3.4.1 Test Dataset Selection and Composition

A stratified random sampling approach will be delivered to select 5-10% of the total dataset (approximately 880 cases from the initial dataset of 8800 children) to serve exclusively as the blind test set. This test cohort will maintain balanced representation with 440 normal weight and 440 overweight/obese subjects, ensuring adequate statistical power for evaluation while preventing data leakage into the training process. The stratification will preserve the demographic and clinical characteristics distribution of the original dataset, maintaining representation across all participating clinical sites.

3.4.2 Expert Evaluation Panels

Clinical Expert Panel

The Clinical Expert Panel will validate the practical utility of the prediction models. The evaluation will focus on clinical relevance of predictions, alignment with established medical knowledge, practical utility in healthcare settings and potential impact on treatment decisions. The main activities will include:

- Clinical assessments on test cases without knowledge of AI predictions
- Evaluating whether identified risk factors align with established obesity pathways
- Assessing the clinical meaningfulness of AI-generated explanations

© 2023-2025 BIO-STREAMS

- Validating the real-world applicability of model outputs in patient care

Each of the seven clinical sites will nominate 2 internal clinicians (already participating in the project) and 2 external clinicians (independent from the project), resulting in approximately 30 clinical evaluators.

Research Expert Panel

The Research Expert Panel will carry out scientific validation focusing on methodological rigor and technical aspects. The evaluation will focus on scientific validity of model construction, statistical performance, reliability across demographic subgroups and model explainability and transparency. The main activities will include:

- Evaluating technical robustness of the ML/AI models
- Analyzing statistical validity of predictions against established metrics
- Assessing explainability frameworks (SHAP, LIME) from a scientific perspective
- Validating methodological approaches in model development

Similarly, pilot sites will provide 2 internal and 2 external researchers each, creating a parallel evaluation team of approximately 30 research experts.

3.4.3 Experiment Workflow

Pre-test preparation will include preparation of pseudonymized test datasets, harmonization of datasets, description of the standardized evaluation forms and metrics capture instruments and training of the evaluators on assessment protocols without revealing model architectures.

The Evaluation will be carried in 4 phases:

(1) Clinical Expert Assessment - Clinicians receive pseudonymized patient cases from the test dataset (440 normal weight, 440 overweight/obese), 30 cases each. Each case includes standardized clinical, demographic, and behavioral data, all harmonized according to BIO-STREAMS Common Semantic Data Model. Evaluators conduct a clinical assessment without any knowledge of AI predictions. Assessments include obesity risk evaluation, identification of key risk factors, and potential intervention recommendations. Structured evaluation forms capture clinical judgments using standardized metrics (see example in Appendix E).
(2) ML/AI Model Predictions - The same 880 pseudonymized test cases are processed through the BIO-STREAMS AI/ML pipeline. Models generate predictions without any exposure to the clinical expert assessments. Predictions include obesity risk scores, influential factor identification, and explainability outputs (SHAP/LIME). Multiple model architectures will be used (including Support Vector Machines, Boosting ensembles, Neural Networks).
(3) Independent Comparative Analysis: Independent data analysis team not involved in model development or clinical assessment carries out Comparative Analysis. Output is collected from both expert and machine assessment without revealing the source of decision. The team will carry out the statistical comparison of clinical vs. machine assessments and calculate performance metrics (sensitivity: 89.18%-90.77%, specificity: 88.93%-91.00%). The ground truth for each case will be the clinical diagnosis (or not having a diagnosis). The team will also qualitatively evaluate concordance/discordance patterns between clinical and AI assessments.
(4) Clinical Research Expert Evaluation: Researchers will receive ML/AI model architectures, training protocols, and anonymized prediction outputs. The team will carry out a systematic evaluation of methodological rigor, and detailed analysis of

© 2023-2025 BIO-STREAMS

model interpretability and explainability mechanisms using standardized assessment framework (see example in Appendix F).

.

© 2023-2025 BIO-STREAMS

4 Recruitment Strategy

In the BIO-STREAMS project we will strategically explore internal resources. Centers will conduct reviews of their clinical departments to identify clinicians with relevant expertise in pediatric care, pediatric obesity, endocrinology, and nutrition, not involved in the BIO-STREAMS project. Department heads will serve as recruitment facilitators, personally engaging potential participants through established professional relationships. To recruit research evaluators the academic departments will identify researchers with methodological expertise and statistical knowledge. Moreover, internal communication systems will be exploited to distribute recruitment materials and invitations through existing clinical networks, staff meetings, and institutional newsletters.

To contact experts outside the institution, centers will analyze the electronic medical records systems, especially focused on referrals, to identify clinicians and general practitioners actively treating childhood obesity cases. Centers will also utilize existing research databases with established communication channels to advertise participation opportunities. In this way local clinical and academic institutions will be contacted to enhance recruitment. Mechanisms such as: primary care connections, health system integration, academic networks, specialty associations and community programs will create a robust framework ensuring sufficient recruitment of participants.

The recruitment process and success will be monitored on a weekly basis, until all the participants identify and screen at least 16 subjects.

© 2023-2025 BIO-STREAMS

5 Regulatory and Ethical Considerations

5.1 Ethics Approvals

All Clinical sites secured ethics approval. The following table summarizes the approvals per site.

Table 2: Ethics Committee Approvals for BIO-STREAMS Project's Retrospective Study

Approval Date	**Ethics Committee**	**Location** and **Contact**	**Reference**
25/04/2024	Ethics Committee of the Specialized Hospital for Rehabilitation and Long-Term Treatment BLOCKS (SBRPL BLOCKS)	Konstantin Pomyanov St., 1, Sofia, 1415, Bulgaria +359 (0)888061383 research@blocks.care	BLOCKS-2024-001
12/03/2024	Ethics Committee of the CHU Liege (707)	Avenue de l'Hôpital, 1, Liege, 4000, Belgium +32 (0) 4.323.00.00 ethique@chuliege.be	2024/82
22/05/2024	Swedish Ethical Review Authority	Box 2110, Uppsala, 750 02, Sweden +46 (0)10-475 08 00 registrator@etikprovning.se	2024-02377-01
01/05/2024	Scientific Council of Athens Children's General Hospital "Agia Sofia"	Thivon & Papadiamantopoulou, Athens, 11522, Greece +30 (0)2132013099 a.tsola@paidon-agiasofia.gr	15593/01-03-2024
14/03/2024	Scientific Council of Children's Hospital PENTELI	Hippocrates 8, Penteli, 15236, Greece +30 (0)213-2052315 esgnpp@paidon-pentelis.gr	3143/12-03-24
28/03/2024	Medical Ethics Committee of University Medical Center Maribor	Ljubljanska cesta 5, Maribor, 2000, Slovenia +386 (0)23212489 eticna.komisija@ukc-mb.si	UKC-MB-KME-17/24
30/05/2024	Clinical Research Ethics Committee of University Hospital VALL D'HEBRON	Paseo de la Vall d'Hebron, 119-129, Barcelona, 08035, Spain	(AG) 127/2024

© 2023-2025 BIO-STREAMS

		+34 (0)934 89 30 00 ceic@vhir.org	

5.2 Informed Consent

The informed consent process will be carried out during the onboarding of subjects. During the process a dialogue with participants will be opened. The recruitment staff of each clinical site will explain the nature of blinding, including which information will be concealed and why. They will also clearly describe circumstances under which unblinding might occur and participants' rights to withdraw consent at any time.

A letter of consent to be further adapted and translated into the local language of the clinical site is given in Appendix B.

5.3 Data Oversight

The study plans to process special categories of personal data (i.e. 'sensitive data') of children, which demands a strong emphasis on data oversight. To ensure compliance with ethical and legal obligations including the General Data Protection Regulation (GDPR), and to safeguard patient privacy, appropriate and state-of-the-art organizational and technical measures will be adopted. This includes the use of encryption technologies to ensure access to information about the study participants is limited to what is necessary to lawfully achieve the purposes of the study. All data processing will occur in the EU.

The data that is collected for the purposes of the study are included in participants' personal medical records and are stored in the same way as other existing medical data. The lawful basis for processing personal data will rely on the research exemption in Article 89 of the GDPR, in combination with either the purpose limitation exemption in Article 5(1)(b) or the lawful basis of public interest necessity (scientific research) in Articles 6(1)(e) and 9(1)(j). Consent cannot be used as a lawful basis because it is impossible to gather specific consent from thousands of subjects for participation in this study due to time, cost, and risk of drop-out.

In accordance with the GDPR's research exemption, appropriate safeguards will be adopted to protect personal data throughout the lifecycle of the study. These measures will be presented during the ethical approval process at each participating institution. The investigators will transfer the structured and pseudonymized data required for the study by applying pseudonymization tools to the data contributed by members of the BIO-STREAMS consortium who will use them. Relevant Data Transfer (DTA) and Data Processing (DPA) Agreements will be signed among the parties involved. Documents collected for the purpose of this clinical investigation will be kept by the consortium until the end of the BIO-STREAMS project. The example DTA/DPA is provided in Appendix C.

5.4 Continued Monitoring and Risk Management

Continued ethical monitoring in clinical studies will involve oversight throughout the entire research lifecycle to ensure participant protection and data integrity remain prioritized from initial approval through study completion. For the BIO-STREAMS study, this process will be carried out according to the Standard Operating Procedures for the Retrospective Study, details are defined in Appendix D.

© 2023-2025 BIO-STREAMS

The BIO-STREAMS consortium has identified the following key risks in the retrospective study:

Table 3: Key Risks Requiring Continued Ethical Monitoring in BIO-STREAMS Study

Risk Category	Specific Risk	Description
Data Security	Unauthorized access	Access attempts by internal or external parties without proper authorization
	Accidental data exposure	Unintentional release of confidential information to unauthorized parties
	Data loss	Potential loss of research data due to technical failures or human error
	Security incidents	Breaches requiring immediate response according to escalation protocols
Privacy & Compliance	GDPR violations	Non-compliance with data protection regulations resulting in regulatory penalties
	Harm to participants	Potential negative impacts on individuals if their sensitive data is compromised
	Protocol deviations	Departures from approved methods affecting data integrity and research validity
	Consent violations	Issues with processing of consent status or withdrawal requests
Operational	Data quality issues	Problems with completeness, accuracy, or consistency before transfer
	Transfer security failures	Vulnerabilities in encryption protocols or authentication mechanisms
	Permission management gaps	Inappropriate access levels or unauthorized role assignments
	Inadequate monitoring	Failures in tracking systems that should detect suspicious activities
Reporting & Oversight	Documentation failures	Incomplete or inaccurate record-keeping of incidents or monitoring activities
	Delayed incident response	Failure to meet the 24-hour emergency response requirement
	Inadequate follow-up	Insufficient post-incident analysis and preventive measure implementation

© 2023-2025 BIO-STREAMS

	Audit trail gaps	Incomplete logging of access, processing, or security events

© 2023-2025 BIO-STREAMS

Conclusions

The BIO-STREAMS retrospective study represents an important first phase in creating a comprehensive EU-wide childhood obesity data space. The observational multi-center cohort study will collect and analyze existing medical records from seven hospitals across six European countries, enabling the development of AI models for obesity risk assessment and providing the foundation for subsequent prospective research.

The BIO-STREAMS study initiation package defines a mixed-methods study to address childhood obesity through two interconnected components: co-creation workshops and a longitudinal intervention study. All data will be collected from existing medical records following strict data protection standards. The records will be pseudonymized at source before being integrated into the BIO-STREAMS Virtual Biobank. Harmonization procedures will ensure consistency across different clinical sites and data formats. The collected data will be used to train initial AI models for risk assessment, which will be validated through randomized blind tests involving 30 clinicians and 30 researchers. This validation will occur in two phases: first with the Initial AI Models (October 2025) and then with the Improved AI Models (October 2026).

All participating sites have received ethical approval from their respective ethics committees. The study employs a federated data management architecture that prioritizes data privacy while enabling collaborative research. No directly identifiable patient information will be shared outside the originating clinical sites.

The study is coordinated by the University of Maribor and the BIO-STREAMS ethics board. Each participating hospital has a designated Principal Investigator who will oversee data collection, study execution and ensure protocol adherence at their respective site.

Co-funded by the European Union

© 2023-2025 BIO-STREAMS

6 Appendix A: Detailed Study Protocol

EU-Wide Childhood Obesity Biobank and Decision Support System: An Observational Cohort Study

Clinical Study Protocol	
Clinical protocol number	1
Document version	2.2
Clinical investigation Title	EU-Wide Childhood Obesity Biobank and Decision Support System: An Observational Cohort Study
Novel solution	New technology – BIOSTREAMS – Childhood Obesity Platform: Knowledge base and Predictive/Prognostic AI systems
Organization responsible for clinical investigation (S)	
Principal Investigator of the Pilot Center	
Coordinating investigator (CI)	Dr. Izidor Mlakar, University of Maribor

© 2023-2025 BIO-STREAMS

Document History

Revision	Date of enactment	Change author	Change description
1.0	07.12.2023	Izidor Mlakar	First draft of the document
1.1	08.01.2024	All partners	First set of comments and feedback
2.0	17.01.2024	Izidor Mlakar	Second draft of the document
2.1	30.01.2024	All partners	Final contributions to content comments and feedback
2.2	05.02.2024	All partners	Final review and production ready protocol

© 2023-2025 BIO-STREAMS

Executive Summary

Title	EU-Wide Childhood Obesity Biobank and Decision Support System: An Observational Cohort Study
Aim	The purpose of this study is to generate a homogenized dataset consisting of retrospective data on demographic, medical (healthcare indices), behavioral and family history information derived from the Electronic Health Records of the clinical sites to (1) identify biological pathways conferring childhood obesity and (2) to validate the predictive/prognostic power of AI algorithms developed by BIO-STREAMS consortium.
Retrospective study design (RSD)	Structured and pseudonymized retrospective data collected from 7 clinical sites. Blind tests to validate theAI algorithms to identify susceptibility and possible health outcomes.
RSD Primary objective	To create a holistic EU Obesity Biobank by combining retrospective data from multiple EU cohorts and allow for further analyses regarding risk factors and biomarkers of overweight in a large sample. To determine the best features as input to the AI models for predicting susceptibility and possible health outcomes.
RSD Secondary objective	The secondary aim of the study is to assess clinician feasibility of integrating the EU Childhood Obesity Platform into regular clinical routine.
RSD Primary endpoint	(3) Number of new obesity pathways discovered (4) Sensitivity and specificity of the AI models for prediction of risks for obesity (compared to the detailed clinical assessment).
RSD Secondary endpoint	Suitability of the EU Childhood Obesity Platform measured through: usability, acceptance, user experience, trust (related to AI predictions) Homogenized retrospective BIO-STREAMS dataset Data curation, data management solution, and infrastructure Cost and economic benefit
Duration of clinical investigation	36 months
Subject population	In total, more than 8800 subjects (5600 children with overweight/obesity 3200 children with normal weight) are estimated to be included across all the study's 7 clinical sites: 8. University Medical Centre Maribor (UKCM) 9. National and Kapodistrian University of Athens (NKUA) 10. Karolinska Institute (KI) 11. Blocks Health and Social Care EOOD (BLOCKS) 12. Hospital Universitari Vall d'Hebron (VHIR) 13. Centre Hospitalier Universitaire de Liège (CHUL) 14. Penteli General Childrens' Hospital (PENTELI)

© 2023-2025 BIO-STREAMS

	Sex distribution: In 2020, around 175 million children and adolescents aged 5-19 years worldwide were considered obese[15]. In the study female and male subjects will be recruited in proportions. However, it must be noted that boys are more susceptible to obesity than girls. A significant difference between genders in European cohorts was in favor of girls, i.e. fewer girls were obese than boys. On average across 23 EU countries, 14% of boys and 10% of girls aged 7-8 years were obese, according to the COSI study[16]. Overall, the prevalence of overweight and obesity among boys is 31%, while among girls it is 28%[17]. The data points of the following subject groups will be included in this retrospective study: Children and adolescents, age of participants: 5-18 years Children and adolescents with normal weight: BMI less than 1 SD above the median of the WHO growth reference for children and adolescents[18], not diagnosed with severe mental disorder, epilepsy , severe chronic medical conditions, cancer, orthopedic problems, known family predisposition, eating disorders or do not have any serious chronic disease. Children and adolescents with Increased Body Weight (overweight) or with Obesity: overweight defined as BMI more than 1 SD to 2 SD above the median of the WHO growth reference for children and adolescents. Obesity defined as BMI more than 2 SD above the median of the WHO growth reference for children and adolescents not diagnosed with an eating disorder, severe mental disorder, epilepsy or severe chronic medical conditions, cancer prior to overweight/obesity. To evaluate primary outcome, sensitivity and specificity data from 5% randomly selected retrospective subjects will be reserved for the validation of the models using blind tests. Experts, 30 healthcare experts and 30 researchers will participate. The experts need to be healthcare professionals involved in clinical research or practice obesity in the pediatric population. The researcher must be engaged in clinical research on Big Data and/or obesity (e.g. epidemiology of obesity, obesity research, etc.).
Number of subjects	In total, more than 8800 subjects are necessary to be included from 7 clinical sites to account for the minimum sample size per group (sample size justification) and the relatively low incidence rate of obesity (< 6%) (UKCM, NKUA, KI, BLOCKS, VHIR, CHUL, PENTELI).

15 https://www.statista.com/statistics/1386146/number-of-obese-children-and-adolescents-worldwide-forecasts-by-gender/

16 *https://www.oecd-ilibrary.org/sites/health_glance_eur-2018-26-en/index.html?itemId=/content/component/health_glance_eur-2018-26-en*

17 *https://www.who.int/europe/news/item/03-03-2023-childhood-obesity--five-facts-about-the-who-european-region*

18 WHO growth reference for children and adolescent: https://www.who.int/tools/growth-reference-data-for-5to19-years

© 2023-2025 BIO-STREAMS

Number of Sites	7 clinical sites: 8) UKCM, 9) NKUA, 10) KI, 11) BLOCKS, 12) VHIR, 13) CHUL, 14) PENTELI
External organizations involved in the clinical investigation	No organizations external to the BIO-STREAMS consortium will be involved in the clinical investigation. Each clinical site, contributing with data, will allocate a pilot leader (e.g. the Principal Investigators) to oversee the operations within the pilot study. Project's Clinical Manager, Izidor Mlakar (University of Maribor) will oversee the execution of all the studies. Oversight of the Data Operations will be carried out by Magdalena Góralczyk, Peter Davis and Lucrezia Nicosia (White Label Consultancy APS) and the Project's Data Protection Officer, Dimitris Kalogeras (ICCS). Each clinical partner also has their own DPO (contact details of each provided below).
Inclusion criteria	▪ Age of participants between 5 and 18 years ▪ (Overweight Cohort): BMI more than 1 SD to 2 SD above the median of the WHO growth reference for children and/or adolescents, ▪ (Obese Cohort): BMI more than 2 SD above the median of the WHO growth reference for children and adolescents ▪ (Normal Weight Cohort): BMI up to 1 SD above the median of the WHO growth reference for children and adolescents[19]
Exclusion criteria	▪ Age under 5 years or above 18 years ▪ BMI lower than 2 SD below the median of the WHO growth reference for children and adolescents ▪ diagnosed with an eating disorder, severe mental disorder, epilepsy or severe chronic medical conditions, cancer prior to overweight/obesity, untreated hypothyroidism
Data Oversight	The study plans to process special categories of personal data (i.e. 'sensitive data') of children, which demands a strong emphasis on data oversight. To ensure compliance with ethical and legal obligations including the General Data Protection Regulation (GDPR), and to safeguard patient privacy, appropriate and state-of-the-art organizational and technical measures will be adopted. This includes the use of encryption technologies to ensure access to information about the study participants is limited to what is necessary to lawfully achieve the purposes of the study. All data processing will occur in the EU. The data that is collected for the purposes of the study are included in participants' personal medical records and are stored in the same way as other existing medical data. The lawful basis for processing personal data will rely on the research exemption in Article 89 of the GDPR, in

[19] WHO growth reference for children and adolescent: https://www.who.int/tools/growth-reference-data-for-5to19-years

© 2023-2025 BIO-STREAMS

	combination with either the purpose limitation exemption in Article 5(1)(b) or the lawful basis of public interest necessity (scientific research) in Articles 6(1)(e) and 9(1)(j). Consent cannot be used as a lawful basis because it is impossible to gather specific consent from thousands of subjects for participation in this study due to time, cost, and risk of drop-out. In accordance with the GDPR's research exemption, appropriate safeguards will be adopted to protect personal data throughout the lifecycle of the study. These measures will be presented during the ethical approval process at each participating institution. The investigators will transfer the structured and pseudonymized data required for the study by applying pseudonymization tools to the data contributed by members of the BIO-STREAMS consortium who will use them. Relevant Data Transfer (DTA) and Data Processing (DPA) Agreements will be signed among parties involved. Documents collected for the purpose of this clinical investigation will be kept by the consortium until the end of the BIO-STREAMS project.
Clinical investigation financing	This Clinical study is part of the BIO-STREAMS European Project funded within the research and innovation program of the Horizon Europe under N° 101080718. The funding source had no impact on the decision to carry out the study or on its design. The funding source will have no impact on the decisions related to publishing this research or its outcomes. The content of this document does not reflect the official opinion of the European Union or any other institution.
Foreseen compensation for costs and time incurred in participating in a clinical investigation, procedure and conditions for calculation and payment of compensation	No compensation is provided.

6.1 General Information

6.1.1 Study Rationale

The rising prevalence of obesity and related metabolic diseases has become a major public health concern worldwide. Childhood obesity is the most prevalent chronic disease in developed and developing countries. The increasing prevalence of obesity in childhood and adolescence is particularly alarming with significant psychological and psychosocial impact, along with the increased risks of adulthood obesity and non-communicable diseases (NCDs) that increase lifetime excess direct costs. Given the epidemic proportions of obesity worldwide and the associated significant morbidity and mortality, it is important to undertake preventive measures early in life.

Social determining factors, such as the adverse influence of globalization, supermarket growth, fast unplanned urbanization, sedentary lifestyle, economy, and social position slowly develop behavioral risk factors in humans. Behavioral risk factors, such as unhealthy lifestyle habits, inappropriate diet, and physical inactivity lead to physiological risks and obesity or overweight. The increase of BMI, especially in adolescence, is associated with a risk of metabolic

© 2023-2025 BIO-STREAMS

disorders, such as type 2 diabetes, metabolic syndrome and fatty liver disease but also other comorbidities such as orthopedic complications, obstructive sleep apnea syndrome, asthma, coronary heart disease, as well as depression and low self-esteem.

The analysis of existing obesity/overweight related data using machine learning algorithms did not produce any brand-new risk factors, but it helped us understand: (a) how risk factors related to weight change are identified and how we detect them (b) what the nature of the data (potential risk factors that can be monitored) to be collected over time would be, to develop personalized behavior modification coaches and minimize/optimize the use of more effective pharmacological interventions and bariatric surgery as much as possible. Furthermore, prevention should begin at early ages to consolidate healthy lifestyle habits, and those who suffer from any of these disorders should seek the help and advice of medical staff. Thus, early detection and evidence of risks are essential.

With increasing volumes and greater access to data in electronic formats, it is unsurprising that researchers are beginning to apply big data to key concerns including mental health[20], infectious disease[21] and healthcare[22]. In the field of obesity research, there is a long history of using routine data sources to track the prevalence of the disease, as well as identify risk factors. Supplementing this with new forms of data has potential to broaden our understanding of obesity, bringing together information from different facets of the environment and behaviors[23]. BIO-STREAMS aims to deliver a Multi-Pillar Framework to address childhood obesity by building on an EU biobank, micro-moments and mobile recommendation systems. To this end, BIO-STREAMS mobilizes a diverse group of partners with clear in-project duties, to design, create, and deploy the following in multiple settings: involving 7 hospitals in 6 EU countries and 5 school sites in 5 EU countries: The first EU Childhood/Adolescence Obesity Biobank (EU data space – BIO-STREAMS Biobank) acting as an EU-wide data-sharing center for research and innovation, hosting real and synthetic data and ensuring:

- Standardized data collection, exploiting knowledge from European Core Health Indicators.
- Data model with demographic, behavioral, clinical, genetic/epigenetic and cost data
- Expandable Data Network hosting diverse datasets across countries via EU-wide local hubs.
- Knowledge Hub with group-specific policies (for health professionals, schools and citizens), best practices and research outcomes, regularly updated upon entry and analysis of new data within the BIO-STREAMS Biobank.
- Personalized Risk-Assessment of adverse metabolic outcomes attributed to obesity.
- Recommendation Engine generating tailored programs for prevention and healthy living, following a family centric approach and considering micro-moments as determinants driving behavior and adherence.
- Evidence-based knowledge communication to stakeholders via transparent methods for analysis & reporting.

---

[20] *Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. Soc Psychiatry Psychiatr Epidemiol. 2016;51:1055–72.*

[21] *Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease. PLoS Med. 2013;10:e1001413*

[22] *Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2:3.*

[23] *Timmins, K. A., Green, M. A., Radley, D., Morris, M. A., & Pearce, J. (2018). How has big data contributed to obesity research? A review of the literature. International journal of obesity, 42(12), 1951-1962.*

© 2023-2025 BIO-STREAMS

6.1.2 Current Knowledge and Background

Overweight and obesity are the leading lifestyle-related causes of clinical and public health concerns. Change in health behavior is central in obesity management. Bodyweight is influenced by genetic, metabolic, behavioral, environmental, cultural and socio-economic influences[24]. The etiology of obesity can be observed from a clinical and public health perspective. The clinical perspective is based on individual variations (genetic and biologic variations), while the public health perspective is based on calorie intake and energy expenditure[7]. Despite major evidence of the important role of environmental factors, such as sedentary lifestyle combined with intake of energy dense nutrition and reduced energy expenditure, there is no doubt for a strong genetic basis of common obesity. Over the last two decades, efforts in identifying and replicating genetic variants predisposing individuals to common forms of obesity were largely characterized by slow progress and limited success, in sharp contrast to the successful gene identification in monogenic and syndromic forms of obesity[25]. Genome-Wide Association Studies (GWAS) revealed important novel insights into genetics of obesity, however, a major limitation in understanding the genetic contribution is owed to the large proportion of unexplained variability of BMI, as identified Single Nucleotide Polymorphism (SNP) markers collectively explain less than 3%–5% of the observed variability[26]. Furthermore, in addition to genetics, it is also crucial to achieving a sustainable balance between energy output and input. The concept of energy balance is pivotal in the context of weight management and obesity. When an individual is in energy balance, meaning energy intake equals energy expenditure, then weight should remain stable[27]. To achieve a sustainable balance, it is important to increase energy expenditure through physical activity and reduce sedentary behavior.

Overall, overweight and obesity result mainly from a sedentary lifestyle and a negative balance between energy consumption and expenditure[7]. The challenge of choosing which feature interactions to study is a barrier to gaining new insights into the principles of how different exposures in the exposome interact to shape health. This is because the traditional prediction tools often lack consideration of nonlinear, collinear and interactive effects among factors[28]. A wide panel of machine-learning and deep-learning models has been developed to facilitate the characterization of these effects[29]. A and Machine Learning (ML) represent a promising pathway towards timely and accurate screening and diagnosis. Namely, AI (i.e. non-knowledge-based systems) can represent a unique opportunity to address the heterogeneity in the pathophysiology of mental illness and diagnosis by harnessing big data and discover new complex patterns in high-dimensional data well beyond human performance[30]. In fact, medical support based on AI is the modern way to tackle complex problems faced in the medical field. AI provides smart ways to manage medical demand. There have been various studies conducted in academic literature which gives us an idea of the importance of AI in the healthcare sector in enhancing the service quality. AI can improve medical treatment or

---

[24] *Behavioural interventions for preventing and treating obesity in adults. Sharma M. Obes Rev. 2007;8:441–449*

[25] Rohde, K., Keller, M., la Cour Poulsen, L., Blüher, M., Kovacs, P., & Böttcher, Y. (2019). Genetics and epigenetics in obesity. *Metabolism*, *92*, 37-50.

[26] Yengo, Loic, Julia Sidorenko, Kathryn E. Kemper, Zhili Zheng, Andrew R. Wood, Michael N. Weedon, Timothy M. Frayling et al. "Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry." *Human molecular genetics* 27, no. 20 (2018): 3641-3649.

[27] *Hill, J. O., Wyatt, H. R., & Peters, J. C. (2013). The importance of energy balance. European endocrinology, 9(2), 111.*

[28] *P. Doupe, J. Faghmous, S. Basu, Machine learning for health services researchers. Value Health 22, 808–815 (2019)*

[29] *C. Colmenarejo. Machine learning models to predict childhood and adolescent obesity: a review. Nutrients 12, 2466 (2020)*

[30] *Fazi MB. Beyond Human: Deep Learning, Explainability and Representation. Theory, Culture and Society; 38. Epub ahead of print 2021. DOI: 10.1177/0263276420966386.*

© 2023-2025 BIO-STREAMS

diagnosis, e.g., HIV drug resistance prediction[31], breast cancer prediction[32], or type 2 diabetes mellitus (T2DM)[33], however, these models cannot be used in an automated and unsupervised manner. In healthcare, where mistakes can cost human life, the unexplainable (i.e., black-box) nature of AI makes it less acceptable for clinicians and regulators[34].

There are 445 clinical studies registered on ClinicalTrials.gov[35] addressing the use of clinical decision support systems in multiple diseases. Only 9 of them are related specifically to obesity[36]. Most of the studies are observational and measure sensitivity and specificity as their main outcome. In the domain of obesity, similar studies are both interventional with a specific focus on weight loss[37]. None of the studies on clinical decision support systems for obesity specifically address the AI-supported design of intervention or creation of new evidence-based guidelines. Only 2 studies address Perceptions and Attitudes[38] and no studies address trust and acceptance[39]. Only 1 study addresses prediction in terms of weight gain[40]. 2 studies on ISRCTN (International Standard Randomised Controlled Trial Number)[41,42] address prediction of obesity again focused on weight gain.

6.1.3 Research Questions & Hypothesis

The main hypothesis of this study is that there are sets of clinical, biochemical, and lifestyle/behavioral variables that can be associated with the different risks and health outcomes of obesity/overweight in children.

To investigate the hypothesis, we define two primary research questions:

Q1: “What are the clinical, biochemical, and lifestyle/behavioral predictors of obesity?”

Q2: “Can the different interactions create new cohorts underlining more efficient and patient-centered interventions?”

---

[31] Riemenschneider, M., Hummel, T., & Heider, D. (2016). SHIVA-a web application for drug resistance and tropism testing in HIV. *BMC bioinformatics*, *17*(1), 1-6.

[32] Montazeri, M., Montazeri, M., Montazeri, M., & Beigzadeh, A. (2016). Machine learning models in breast cancer survival prediction. *Technology and Health Care*, *24*(1), 31-42.

[33] Talaei-Khoei, A., & Wilson, J. M. (2018). Identifying people at risk of developing type 2 diabetes: a comparison of predictive analytics techniques and predictor variables. *International journal of medical informatics*, *119*, 22-38.

[34] *Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA - Journal of the American Medical Association; 320. Epub ahead of print 2018. DOI: 10.1001/jama.2018.17163.*

[35] *https://clinicaltrials.gov/ct2/results?term=clinical+decision+support+system*

[36] *https://clinicaltrials.gov/ct2/results?cond=Obesity&term=clinical+decision+support+system&cntry=&state=&city=&dist=&Search=Search*

[37] *https://clinicaltrials.gov/ct2/show/NCT03404999?term=clinical+decision+support+system&cond=Obesity&draw=2&rank=6*

[38] *https://clinicaltrials.gov/ct2/results?cond=Obesity&term=clinical+decision+support+system+Perceptions+and+Attitudes&cntry=&state=&city=&dist=&Search=Search*

[39] *https://clinicaltrials.gov/ct2/results?cond=Obesity&term=clinical+decision+support+system+trust+acceptance&cntry=&state=&city=&dist=&Search=Search*

[40] *https://clinicaltrials.gov/ct2/show/NCT01111812?term=prediction&type=Obsr&cond=Obesity&age=0&draw=2&rank=1*

[41] *https://www.isrctn.com/ISRCTN24468014?q=prediction&filters=condition:Obesity&sort=&offset=2&totalResults=2&page=1&pageSize=10*

[42] *https://www.isrctn.com/ISRCTN59323751?q=prediction&filters=condition:Obesity&sort=&offset=1&totalResults=2&page=1&pageSize=10*

© 2023-2025 BIO-STREAMS

The secondary research questions are defined to assess the feasibility and usability of BIO-STREAMS EU Childhood Obesity Platform, and the ML/AI models built based on Q1 and Q2 in clinical practice. To this end we define two secondary research questions:

Q3: "How do clinicians evaluate the feasibility of using the system in clinical routine, based on usability, trust and acceptance, and user experience while using EU Childhood Obesity Platform?"

Q4: "What is the cost and economic benefit and the suitability of the EU Childhood Obesity Platform from the perspective of healthcare systems and policy makers?"

6.1.4 Primary and Secondary Aims Outcomes

Primary aims:

A1: To create a holistic EU Obesity Biobank by combining retrospective data from multiple EU cohorts that will allow further analyses regarding risk factors and biomarkers of overweight in a large sample.

A2: To discover new sub-cohorts of obesity and new biological pathways conferring susceptibility based on clinical, environmental, socio-economic, and lifestyle interaction.

A3: To validate the predictive power of AI algorithms in EU Obesity platform

Primary outcomes:

Number of new obesity pathways discovered.

Sensitivity and specificity of the Machine/Deep Learning models for prediction of risks for obesity (compared to the detailed clinical assessment).

Secondary aim:

A4: The secondary aim of the study is to assess clinician feasibility of integrating the EU Childhood Obesity Platform into regular clinical routine.

Secondary outcomes:

- Suitability of the EU Childhood Obesity Platform measured through usability, acceptance, user experience and trust (related to AI predictions),
- Data curation, data management solution, and infrastructure
- Homogenized retrospective BIO-STREAMS dataset
- Cost and economic benefit

6.1.5 Study Design

Observational, Retrospective Cohort, Randomized, Blind Tests

Based on this study, we will identify risk factors associated with the development of obesity and obesity comorbidities. We will also investigate how these risk factors can be used to prognose health related outcomes. Data from the Electronic Health Records will be utilized and analyzed to extract variables defined as risk factors based on the normal weight, overweight and obese cohorts. Various methodologies, including statistical analysis and machine learning techniques, will be employed to investigate the influence of each factor on Obesity and its health-related outcomes. No interventions *are* foreseen in this study. The experts will validate and evaluate the BIO-STREAMS Platform, its feasibility and usability and the predictive/prognostic power of the ML/AI algorithms.

© 2023-2025 BIO-STREAMS

6.1.5.1 Sample Size Justification

To evaluate primary outcome, sensitivity and specificity of the prediction models, approaches for testing AI and ML models are chosen[43,44]. Data from 5%-10% randomly selected subjects (depending on the final size and complexity of predictors) will be reserved for testing of the model and will not be used in training or validation of the models. Blind tests will be carried out to compare the results of the AI algorithms compared to the detailed clinical assessment. In total the initial dataset will constitute over 8800 children. We estimate the randomly selected subgroup will consist of 880 cases (440 normal weight and 440 overweight/obese). This will give us a pool of a maximum of 44 predictors (i.e. 10 cases per clinical predictor[45]). Each clinical site is estimated to carry out case studies (comparisons) for 126 cases (63 normal weight and 63 for overweight/obese).

The different risk factors (categorical and numerical) will be tested in terms of their association. In terms of categorical variables, using the chi-squared test, the minimum sample size required per group was calculated using the following assumptions:

- Required power of 0.95
- Significance level: 0.05
- Effect size: 0.1
- Two-sided differences
- Average prevalence of overweight and obesity: 29%
- Dropout rate: 25% (to account for the exclusion of participants' records due to errors or missing key variables)
- Target specificity and sensitivity: 85%

The minimum required sample size to detect differences in categorical risk factors is N = 2474 subjects. AI is based on the analysis of large datasets and requires a continuous supply of high-quality data. Data is critical for AI because it is the foundation upon which machine learning algorithms learn, make predictions, and improve their performance over time. To train an AI model, large amounts of data are required to enable the model to recognize patterns, make predictions, and improve its performance over time[46,47]. Thus, in the study we foresee to include data collected from 5600 obese/overweight subjects and 3200 normal weight subjects. Considering a precision of less than 5% for the calculation of sensitivity and specificity of the AI risk factor model and prevalence of 30%, we can expect sensitivity between 89.18% to 90.77% and specificity between 88.93% to 91.00%. Positive Predictive Value (probability that the disease is present when the test is positive) is estimated at average 79.41% and Negative Predictive Value (probability that the disease is not present when the test is negative) is estimated at average 95.45%.

To evaluate the secondary goal, i.e., the usability of the platform via evaluations of clinicians and researchers (2 groups), each clinical site will ensure at least 2 internal (related to project) and 2 external clinicians. In total, at least 30 clinicians will participate. Similarly, each academic

---

[43] Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of analysis and testing*, *2*(3), 249-262.

[44] Reitermanova, Z. (2010, June). Data splitting. In *WDS* (Vol. 10, pp. 31-36). Prague: Matfyzpress.

[45] Gearing, R. E., Mian, I. A., Barber, J., & Ickowicz, A. (2006). A methodology for conducting retrospective chart review research in child and adolescent psychiatry. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 15(3), 126.

[46] *Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy AI. Nature Machine Intelligence, 4(8), 669-677.*

[47] *Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. Applied Sciences, 13(12), 7082.*

© 2023-2025 BIO-STREAMS

institution will ensure at least 2 internal (related to project) and 2 external researchers. In total, at least 30 researchers will participate.

6.1.5.2 Timing of Study Procedures

The foreseen schedule of the study is as follows: 24 months for creating the EU Childhood Obesity Platform and technologies, 6 months for collection and curation, 12 months for validation, 3 months for fine tuning of the technology, and 6 months for data analysis and synthesis of results. The diagram below outlines the study design. The detailed schedule is defined as follows:

- M03 (July 2023) – M09 (Jan 2024): development of the study protocol (expert meetings, meta/scoping reviews)
- M10 (Feb 2024) – M12 (April 2024): final study design approved by Ethics Committees application; Data size, availability and quality estimation. DTA and DPAs signed; Registration of the study at clinical trials registry, ISRCTN Registry (https://www.isrctn.com/)
- M12 (April 2024) – M23 (March 2025): data extraction and transformation into structured data. Data pseudonymization and Data Transfer;
- M12 (April 2024) – M29 (September 2025): data processing and analysis by technical partners. Development of Technology and Models version 1
- M24 (April 2025) – M29 (September 2025): recruitment of experts for blind tests
- M30 (October 2025) – M35 (March 2026): validation of the technology by domain experts
- M36 (April 2026) – M38 (June 2026): fine-tuning and delivery of Minimal Viable Product (MVP) by technical partners
- M39 (July 2026) – M44 (Dec 2026): validation of the technology by domain experts
- M45 (Jan 2027) – M48 (April 2027): analysis and publication of results

6.1.5.3 Description of Study Procedures

6.1.5.3.1.1 Development of the Study Protocol and Variable List Definition

General clinical preparation and expert meetings have already been held in order to elucidate a more suitable variable list for BIO-STREAMS. The inclusion and exclusion criteria have been modified to be applicable to the study. Demographic data, medical and family history, clinical biology, healthcare indices, behavioral data, and detailed diagnosis of obesity are included in the final list. The final list is adapted based on the results of systematic reviews. Data templates and corresponding encoding instructions are provided to all clinical partners.

6.1.5.3.1.2 Data Size, Availability, and Quality Estimation

According to BIO-STREAMS' inclusion and exclusion criteria, relevant participants' electronic health data are recruited for the retrospective study. The sample size, variable availability, and data quality are reported, and the exact data size will be coordinated within the consortium to have a better data balance per group for BIO-STREAMS. Minimal Initial sample size per both cohorts and pilot site are reported in the following table:

Table 4: The distribution of retrospective cohorts

Partner	Country	Overweight/Obese	Normal Weight
University Clinical Centre Maribor	Slovenia	1100	500

© 2023-2025 BIO-STREAMS

National and Kapodistrian University of Athens	Greece	2500	300
Karolinska Institute	Sweden	250	1550
Blocks Health and Social Care EOOD	Bulgaria	50	50
Hospital Universitari Vall d'Hebron	Spain	1000	300
Centre Hospitalier Universitaire de Liège	Belgium	400	200
Penteli General Children's Hospital	Greece	300	300
Total		**5600**	**3200**

Data Extraction and Transformation into Structured Data

Data will be extracted by pilot sites in a common format defined by WP3 of the project, and unstructured data will be transformed into a unified content and representation form via a data harmonization pipeline tapping into the BIO-STREAMS common data model and ontology. In particular, the BIO-STREAMS dataset (data) and dataset descriptor (metadata) will be built in a standardized format by employing a common data model, e.g, the CDISC[48] or OMOP[49], which comply *per se* with community standard domain models, terminologies and formats (e.g., ICD10, SNOMED, LOINC). In addition, the dataset will be contextualized by being associated with the BIO-STREAMS data ontology, providing domain knowledge, which will enhance semantic interoperability. The BIO-STREAMS common data model, ontology and hosting environment, all together, contribute to assuring a high FAIR dataset maturity level[50,51,52]. However, public sharing of the retrospectively collected dataset beyond the consortium members (i.e. 3rd parties) is not foreseen.

Data Pseudonymization and Transfer

The study plans to process special categories of personal data (i.e. 'sensitive data') of children, which demands a strong emphasis on data oversight. To ensure compliance with ethical and legal obligations including the General Data Protection Regulation (GDPR), and to safeguard patient privacy, appropriate and state-of-the-art organizational and technical measures will be adopted. This includes the use of pseudonymization technologies to ensure access to information about the study participants is limited to what is necessary to lawfully achieve the purposes of the study. All data processing will occur in the EU.

The data that is collected for the purposes of the study are included in participants' personal medical records and are stored in the same way as other existing medical data. The lawful basis for processing personal data will rely on the research exemption in Article 89 of the GDPR, in combination with either the purpose limitation exemption in Article 5(1)(b) or the lawful basis of public interest necessity (scientific research) in Articles 6(1)(e) and 9(1)(j). Consent cannot be used as a lawful basis because it is impossible to gather specific consent from thousands of subjects for participation in this study due to time, cost, and risk of drop-out. Failure to collect consent for certain individuals may also create bias in the outcomes.

In accordance with the research exemption, appropriate safeguards will be adopted to protect personal data throughout the lifecycle of the study. These measures will be presented during the ethical approval process at each clinical site acting as the Data Controller. The investigators will transfer the structured and pseudonymized data (by applying pseudonymization tools offered by the technical team, see paragraph below) required for the

[48] https://www.cdisc.org/standards

[49] https://www.ohdsi.org/data-standardization

[50] Welter, D., Juty, N., Rocca-Serra, P. et al. FAIR in action - a flexible framework to guide FAIRification. Sci Data 10, 291 (2023).https://doi.org/10.1038/s41597-023-02167-2

[51] https://fairplus.github.io/Data-Maturity/

[52] https://faircookbook.elixir-europe.org/content/recipes/maturity.html

© 2023-2025 BIO-STREAMS

study to the members of the BIO-STREAMS consortium who will use them to (i) define and establish common semantic data model and (ii) to train AI/ML algorithms. Relevant Data Transfer (DTA) and Data Processing (DPA) Agreements will be signed among parties involved. Documents collected for the purpose of this clinical investigation will be kept by the consortium until the end of the BIO-STREAMS project.

To ensure the privacy of the data, technical partners will provide pseudonymization tools to the clinical partners in order to pseudonymize the retrospective data. These tools will be utilized by clinical partners. The most appropriate pseudonymization techniques, ranging from basic solutions (e.g., cryptographic hash function, message authentication code, symmetric encryption) to most advanced ones (e.g., Merkle trees, secure multiparty computation, secret sharing schemes) and the pseudonymization policy (i.e., deterministic pseudonymization, document-randomized pseudonymization, fully randomized pseudonymization) will be selected considering the data protection level, the utility of the pseudonymized dataset and the complexity associated to a certain scheme in terms of implementation and scalability[53,54]. Regarding the transferring of the data, an interface will be developed where heterogeneous clinical sites will upload their data in a standardized format. After the harmonization process, the data will be stored inside an external infrastructure, providing API calls retrieval and additional services.

Security measures and data handling procedures of data Processors

Access control: A strong access control (e.g. Attribute-based *access control using SAPL*) will be implemented to limit access to sensitive clinical data, including role-based access control and the principle of least privilege.

Encryption: Encryption technologies, **such as SSL**, will be implemented to protect data in transit and at rest, ensuring that sensitive information is encrypted both when stored and transmitted.

Data usage controls: Data usage controls to monitor and detect unauthorized use of sensitive clinical data, such as web uploads, unauthorized email sends, copying to external drives, or printing, will be implemented.

Logging and monitoring: Comprehensive logs will be maintained and monitoring systems to track user access and activities, enabling prompt detection and response to potential security incidents will be deployed.

Risk assessments: Regular risk assessments to identify potential risks of reidentification and security threats and vulnerabilities will be carried out and appropriate measures to mitigate these risks implemented. The process will be integrated into the project's regular risk monitoring activity.

Data subject rights: Mechanisms will be put in place to ensure that data controllers have processes in place to handle data subject requests, such as access, correction, or deletion of personal data.

Vendor management: A DPA will be signed with data processors of the BIO-STREAMS project to evaluate and manage relationships with data controllers, ensuring that all data processors have the necessary data protection measures in place.

Incident response plans: Incident response plans will be designed to quickly identify, contain, and recover from security incidents involving clinical data.

---

[53] https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases

[54] https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques

© 2023-2025 BIO-STREAMS

Data Analysis and Machine Learning

Rigorous data curation and harmonization pipelines will be employed to assure data quality and standardized data representation, respectively[55,56]. The outcome will be a curated and harmonized dataset, which will be used for risk factor analysis (i.e., predictive modeling). **Data curation** includes: (i) the assessment of data representativeness through a detailed description of the collected data (e.g., the time span of data collection, the collection site and setting, relevant population characteristics such as gender, age, ethnicity, and relevant medical history, and any inclusion or exclusion criteria that were used), aiming at combating selection bias, and, subsequently, algorithmic bias; and (ii) the assessment of data quality through the inspection and description of missing data, outliers and duplicates, considering, in parallel, potential errors in measurement and their underlying mechanisms (e.g., random or systematic), and any known data quality risks and limitations. **Data harmonization** will be employed to address issues related to multiple data sources with varying standards, formats, schemas, structures and ambiguous semantics, and generate a coherent dataset. Data harmonization pipelines, combining lexical and semantic analysis, will draw upon the BIO-STREAMS common data model and data ontology, both aligned with relevant coding standards and widely adopted protocols[57,58] to facilitate data interoperability and reusability[59,60]. **Data pre-processing**, an intermediate layer between data curation/harmonization and predictive modeling, will be employed to handle the identified data quality issues and prepare data for the subsequent phase of predictive modeling. Paradigms of pre-processing steps include removing outliers, deduplication, imputing missing data, transforming variables or creating new features. **Predictive modeling:** Given the BIO-STREAMS feature space, its predictive (prognostic) value with respect to obesity and health outcomes (including comorbidities) will be examined. Machine learning/deep learning algorithms along with statistical analysis will be employed to examine the importance of each feature and identify the most important (i.e., highly ranked) risk factors that affect obesity and health outcomes. Explainable AI methods along with visualization plots will be developed to increase the transparency and interpretability of machine learning and deep learning models.

6.1.5.3.1.3 Data Collection Requirements

Documents collected for the purpose of this clinical investigation will be kept by the Organization responsible for clinical investigation until the BIO-STREAMS project is finished. Data collected during the study will be included in their personal medical records and stored at the Investigator's Office in the same way as other personal medical data and information.

Each Principal Investigator at Each clinical site is required to maintain records of each subject's case history. Source documents include the participant's hospital files (electronic or paper). The Principal Investigator (or members of her/his team) will record which retrospective subject is enrolled in this clinical investigation. Access to the subject records and other source data must be provided to study monitors, auditors, and/or inspectors.

6.1.5.3.1.4 The Use of Retrospective Data

The BIO-STREAMS methodological approach has been separated into distinct implementation Phases. The first Phase involves the retrospective collection, demographic, behavioral, environmental, medical and healthcare indices, and family history data. The first Phase involves the retrospective collection of demographic, behavioral, environmental, medical and

---

[55] de Hond, A.A.H., et al., npj Digit. Med. 5 (2), 2022. https://doi.org/10.1038/s41746-021-00549-7

[56] World Health Organization. (2023). https://iris.who.int/handle/10665/373421. License: CC BY-NC-SA 3.0 IGO.

[57] https://www.cdisc.org/standards

[58] https://www.ohdsi.org/data-standardization

[59] Wilkinson, M., et al. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[60] Rocca-Serra, P., et al. Sci Data 10, 292 (2023). https://doi.org/10.1038/s41597-023-02166-3

© 2023-2025 BIO-STREAMS

family history data as well as healthcare indices. The data to be collected include but is not limited to (subject to availability):

- demographic data: sex, age, education,
- behavioral data: physical activity, sleep duration, eating and drinking patterns,
- family data: BMI and the place of residence
- medical data: subject's BMI (body mass index), BSA (body surface area), birth weight and height-weight in relation to gestational age, breastfeeding duration

The collected data will be explored to reveal correlations with underage obesity and overweight status and subsequently exploited to identify risk factors and prognostic biomarkers of underage overweight/obesity, incorporating ML risk stratification algorithms and validating their predictive power.

The clinical partners will provide Electronic Health Records (EHR) in a structured format based on a reference model, which includes terminologies that are linked with standardized methodologies (e.g. SNOMED, LOINC, ICD10). The BIO-STREAMS cohorts will be extended with data from existing EU and worldwide cohorts (e.g. UK Biobank, ECLS-K datasets). For the pseudonymization process, technical partners will provide specific tools to clinical sites in order to ensure the pseudonymization of data prior to transfer. During this phase, advanced data curation workflows will be developed to provide 4 core functionalities for effective data quality control: (i) detection of outliers (e.g., z-score, isolation forest, elliptic envelope, local outlier factor), (ii) detection of duplicates (based on correlation and lexical matching), (iii) detection of missing values, and (iv) detection of inconsistencies and errors. Data quality evaluation reports will be generated which will be accompanied by data representativeness metadata, with the latter enabling the assessment of selection bias. The data pre-processing pipeline, including methods such as estimation of missing values, deduplication, feature transformation and dimensionality reduction, will lie between data curation/harmonization and the predictive modeling part of data analysis.

This Phase will result in the BIO-STREAMS homogenized pseudonymized dataset, which will be utilized in the second Phase, where the risk factor analysis will be conducted. On the basis of this dataset, risk factors will be identified to examine their association with obesity and a risk assessment ML tool will be developed. For the risk assessment tool, the selection of statistical and ML methodologies will be critically dependent on the data's specific nature. For statistical analysis, methods such as the chi-squared test and ANOVA will be utilized with a clear rationale for their application. Dimensionality reduction techniques, such as principal component analysis, will be applied, if necessary, to manage high-dimensional data effectively. The choice of supervised and unsupervised ML techniques will also rely on the availability and diversity of the collected data. For example, methods like Fuzzy Clustering, Support Vector Machines, Boosting ensembles, Convolutional Neural Networks, Long-Short Term Memory algorithm, Adversarial Networks and Transformers will be employed to investigate the influence of each factor on obesity. Model evaluation and validation will form a core part of the methodology. Techniques like cross-validation will be used to enhance the generalization of the models, and specific evaluation metrics will be calculated. SHapley Additive exPlanations (SHAP) values, along with representative plots, and Local Interpretable Model-Agnostic Explanations (LIME) will be utilized to increase the transparency and interpretability of machine learning models and provide local and global explanations of the developed models. During this work, a holistic obesity risk factor set, and a set of biomarker 'hits' will be produced to be further evaluated in a subsequent analytics step. In addition, this knowledge will be incorporated into the BIO-STREAMS mobile application, which will monitor information and suggest individualized behavioral interventions. Data management and processing nodes for AI modules will be set up, utilizing the ARIS HPC system with Louros DC as the central node, endorsing EU-based green data centers.

The third Phase consists of the evaluation and validation of decentralized infrastructure and integration of the infrastructure and models in BIO-STREAMS Biobank (See figure 1). The

© 2023-2025 BIO-STREAMS

BIO-STREAMS Biobank overcomes limitations on EU-wide penetration, dataset incorporation and ownership found in standard centralized biobanks by relying on a federated architecture that transforms each data provider into a BIO-STREAMS Biobank Hub with localized storage of pseudonymized and harmonized datasets. All connected Hubs will form a Hub Network corresponding to the BIO-STREAMS Biobank. Specifically, each BIO-STREAMS Member will install on its premises the BIO-STREAMS Node Bundle (BNB) encompassing all necessary infrastructure and services for setting up a Local Hub, including tools for privacy-preserving data management & interaction with BIO-STREAMS main services and other Hubs. BIO-STREAMS will provide technical and operational support regarding installation, deployment and maintenance, enabling data providers to join the Hub Network (i.e. the BIO-STREAMS Biobank). In the third Phase each data controller will deliver the BIO-STREAMS Node Bundle that consists of:

- Computer resources for deployment,
- Database schema & services,
- Regulatory protocols & workflows,
- Data audit & harmonization tools,
- Pseudo & anonymization tools,
- Bio-Streams Platform API
- Federated AI tools

The pseudonymized retrospective data collected within this protocol also assumes a crucial role in the synthetic data generation process. The synthetic generation process is designed to establish a framework that reflects the inherent design, structure, and configurations found in real-world retrospective data particularly in the creation of realistic datasets. BIO-STREAMS will produce custom datasets on-demand for training purposes and fine-tuning of ML pipelines, addressing missing data/metadata, reducing bias and improving reliability. BIO-STREAMS' synthetic datasets will mimic actual data (format and attribute relationships) with no relation to any actual case, consisting of complex imitation versions of actual data rather than modified data. Moreover, use of synthetic datasets will allow for reliable methodology testing on large virtual cohorts, extrapolating from populations of small sample sizes. To implement the synthetic data generator, state-of-the-art deep learning techniques will be employed.

6.1.5.4 Risk-to-benefit Rationale

There are 445 clinical studies registered on ClinicalTrials.gov addressing the use of clinical decision support systems in multiple diseases. Only 9 studies are related specifically to obesity. Most of the studies are observational and measure sensitivity and specificity as their main outcome. In the domain of obesity, similar studies are interventional with a specific focus on weight loss. None of the studies on clinical decision support systems for obesity specifically address the AI supported design of intervention or creation of new evidence-based guidelines. Only 2 studies address Perceptions and Attitudes and no studies address trust and acceptance. Only 1 study addresses prediction in terms of weight gain. 2 studies on ISRCTN, address prediction of obesity again focused on weight gain. None of the studies report on any safety issues related to behavioral interventions.

6.1.5.5 Study Deviation and Changes

A study deviation is an event where the investigator or investigation site personnel did not conduct the clinical study according to the Clinical Study Protocol. The investigator is not allowed to deviate from the above-mentioned documents except with prior approval and under emergency circumstances. All deviations shall be documented and explained, regardless of the reason for the deviation. All major deviations shall be promptly reported to the clinical trial

© 2023-2025 BIO-STREAMS

registry. When necessary, ethics revaluation and an amendment to the ethics approval will be requested.

6.1.6 Study Quality Control Procedures

6.1.6.1 Data Review and Processing

During the study, the completeness of patient records will be checked based on the accuracy of entries, the adherence to the protocol and to Good Clinical Practice as well as GDPR protocols, and the progress of enrolment.

Data management will be carried out according to BIO-STREAMS internal procedures as outlined in the Report on Data Management Policy and the Data Management Plan for this clinical investigation. These documents provide further details as to how BIO-STREAMS partners must process data lawfully and ethically, and can be made available upon request.

All collected data will be reviewed for completeness, correctness, and consistency, in accordance with the Data Management Plan. In case of issues, queries will be sent to the investigator to complete, correct or comment on the data.

1.1.1.1 Study Suspension or early termination

The study may be terminated or suspended at the initiative of the investigators if any of the following reasons arise:

- Data Privacy Concerns: If concerns regarding patient privacy and data protection arise, they may lead to the suspension or termination of the protocol. This could occur in case of breaches in data security, unauthorized access to patient records, or non-compliance with data protection regulations.
- Legal or Regulatory Issues: should legal or regulatory violations related to the study emerge, such as non-compliance with institutional policies, local regulations, or applicable laws, the protocol procedures may be suspended or terminated to address these issues.
- External Factors: External circumstances such as natural disasters, public health emergencies, or unforeseen events that disrupt the healthcare system or impede data access and retrieval from EHRs may necessitate the suspension or termination of the protocol procedures.

In this case, the investigator(s) must inform the Organization responsible for clinical investigation of the reasons for the termination of the study, and the data collected prior to the termination of the study must be passed on to the BIO-STREAMS Platform.

Any changes will be agreed in advance with the Bioethical Committee that authorized the clinical investigation.

6.1.6.2 Study Close Out

Principal investigator of the clinical site and/or its designees will notify the site of the intention to close the study. Study close-out visits may be performed. During these visits, the monitors will ensure that the investigator's regulatory files are up to date and complete and that any outstanding issues from previous visits have been resolved. Principal investigator of the clinical

© 2023-2025 BIO-STREAMS

site and/or its designees will notify and inform the site(s) that all requirements have been met with a study closure letter.

Principal investigator of the clinical site and/or its designees will notify the relevant Bioethical Committee about the clinical trials closure by providing a Clinical Studies report based on the Bioethical Committee/another regulatory authority form.

The Clinical Coordinator of BIO-STREAMS Studies will update the information on the study in the relevant study registry accordingly.

1.1.1.2 Data Reporting and Publication

Any deviations from the CIP will be described and justified in the Final Clinical Study Report, as appropriate.

Publications and presentations referring to this clinical study will be coordinated by Principal Investigators of the Clinical Sites, Izidor Mlakar (acting as the clinical coordinator of BIO-STREAMS Studies) and the Project's Data Protection Officer, Dimitris Kalogeras (ICCS) to ensure privacy preserving use of all available data and confidentiality of shared results.

The study will be registered, and progress and results of the study will be regularly updated at the public clinical study registry, ISRCTN Registry (https://www.isrctn.com/).

6.2 BIO-STREAMS Clinical Sites contributing with data

Research center No 1: University Clinical Centre Maribor

Address: Ljubljanska 5, 2000 Maribor

Title of the department (s): University division of Paediatrics, Head: Jernej Dolinšek, MD, Pediatrician, e-mail: jernej.dolinsek@ukc-mb.si.

DPO: mag. Klara Mihaldinec, e-mail:dpo@ukc-mb.si

Principal investigator: Martin Bigec, MD Pediatrician, e-mail:Martin.Bigec@ukc.mb.si

Research team: Jernej Vidmar, Specialist in Psychology (Jernej.VIDMAR@ukc-mb.si), Kaja Golija (Kaja.GOLIJA@ukc-mb.si), Sonja Golob Jančič, MD (Sonja.GOLOBJANCIC@ukc-mb.si), Vesna Savić (Vesna.SAVIC@ukc-mb.si), Mojca Podgoršek, Nutrition Specialist (Mojca.PODGORSEK@ukc-mb.si), Evgenija Homšak, MD (Evgenija.HOMSAK@ukc-mb.si), Aljaž Valič (aljaz.valic@ukc-mb.si)

Project coordination team: Sergej Černčič (sergej.cerncic@ukc-mb.si), Aljaž Hölbl (aljaz.holbl@ukc-mb.si), Maja Molan (maja.molan@ukc-mb.si), Mojca Hadela (mojca.hadela@ukc-mb.si),

Research center No 2: National and Kapodistrian University of Athens

Address: Thivon and Levadias, 11527, Athens, Greece

Tel: +302132013384.; Fax: N/A; e-mail: childhood-obesity@med.uoa.gr

© 2023-2025 BIO-STREAMS

Title of the department (s): Center for the Prevention and Management of Overweight and Obesity in Childhood and Adolescence, Division of Endocrinology, Metabolism and Diabetes, First Department of Pediatrics, National and Kapodistrian University of Athens Medical School, "Aghia Sophia" Children's Hospital

Tel: +302132013384; Fax: N/A; e-mail:childhood-obesity@med.uoa.gr.

DPO: Mr. Damianos Kosmidis; e-mail: dkosmidis@space.gr & dpo@uoa.gr.

Principal investigator: Professor Evangelia Charmandari (evangelia.charmandari@googlemail.com)

Research team: Professor Evangelia Charmandari (evangelia.charmandari@googlemail.com)

Research center No 3: Karolinska Institute

Address: Alfred Nobels allé 8, 141 52, Stockholm, Sweden

Title of the department (s): Department of Biosciences and Nutrition (BioNut)

DPO: Mats Gustavsson; e-mail: registrator@ki.se.

Principal investigator: Billy Langlet (billy.langlet@ki.se)

Research team: IMPACT research group, Ioannis Ioakeimidis, Billy Langlet, Alkyoni Glibi, Anna Ek

Research center No 4: Blocks Health and Social Care EOOD

Address: 1, Konstantin Pomianov str, 1415 Sofia, Bulgaria

Tel: +359888061383; e-mail: research@blocks.care

Title of the department (s): Phsiotherapy and rehabilitation for children

Tel: +359888061383; e-mail: research@blocks.care .

DPO: Ventsislav Bozhikov; e-mail: v.bozhikov@bozhikov-vatev.bg.

Principal investigator: Assoc. prof. Radka Savova, MD, PhD (savova.radka@gmail.com)

Research team: Assoc. prof. Radka Savova, MD, PhD (savova.radka@gmail.com)

Research center No 5: Hospital Universitari Vall d'Hebron

Address: Pg. de la Vall d'Hebron, 129, Horta-Guinardó, 08035 Barcelona, Spain

Tel: +34 934 89 30 00

Title of the department (s): Vall d'Hebron Institut de Recerca

DPO: Fundació Tic Salut Social, email: dpd@ticsalutsocial.cat.

© 2023-2025 BIO-STREAMS

Principal investigator: Andreea Ciudin. Email: andreea.ciudin@vallhebron.cat

Research team: Andreea Ciudin. Email: andreea.ciudin@vallhebron.cat

Research center No 6: Centre Hospitalier Universitaire de Liège

Address: Avenue de l'Hôpital, 1 4000 Liège, Belgium

Tel: +32(0)4.323.00.00; Fax: N/A; e-mail: *info@chuliege.be*

Title of the department (s): Department of Pediatrics

Tel:+32(0)4.323.92.00; Fax:N/A ; e-mail: N/A

DPO: Mrs Ghislaine Dumont; e-mail: ghislaine.dumont@chuliege.be

Principal investigator: Dr. Caroline Gernay (cgernay@chuliege.be)

Research team: Dr. Caroline Gernay (cgernay@chuliege.be)

Research center No 7: Penteli General Children's Hospital

Address: Ippokratous 8 15236

Tel: +302132052507; e-mail: info@biostreams-penteli.eu.

Title of the department (s):Penteli General Children's Hospital

Tel: +302132052507 e-mail: info@biostreams-penteli.eu.

DPO: Antonia Andriopoulo; e-mail: dpo@paidon-pentelis.gr

Principal investigator: Georgios GkritzelasResearch team: …Olga Fafoula, Athina Balaska, Athanasia Harokopou, Andriana koulountzou, Georgios Feretzakis, Georgios Zagkavieros, Efstathia katoikou , ilias Dalainas

6.3 BIO-STREAMS technical partners involved in data processing

Research partner No 1: University of Maribor

Address: Koroška cesta 46, 2000 Maribor

Title of the department (s): Laboratory for Digital Signal Processing

Principal investigator: …dr. Izidor Mlakar……, e-mail: izidor.mlakar@um.si

© 2023-2025 BIO-STREAMS

DPO: doc. dr. Miha Dvojmoč (dpo@um.si)

Research team: dr. Urška Smrke (urška.smrke@um.si)

Reason for data access: clinical coordination and updates to public study registry, coordination and publication of results, public dissemination

Research partner No 2: National and Kapodistrian University of Athens

Address: Thivon and Levadias, 11527, Athens, Greece

Title of the department (s): Center for the Prevention and Management of Overweight and Obesity in Childhood and Adolescence, Division of Endocrinology, Metabolism and Diabetes, First Department of Pediatrics, National and Kapodistrian University of Athens Medical School, "Aghia Sophia" Children's Hospital

Principal investigator: Professor Evangelia Charmandari, e-mail: evangelia.charmandari@googlemail.com.

Research team: Penio Kassari (peniokassari@gmail.com), Sofia-Maria Genitsaridi (sgenitsaridi@gmail.com), Eleni Ramouzi (eleni_ramouzi@hotmail.gr), Eleni Kokkou (g.elenaki@ymail.com), Marina Papadopoulou (marinageorpap@gmail.com) .

Reason for data access: data collection, structuring, storage, use or disclosure, contribution to the creation of common semantic data model

Research partner No 3: PANEPISTIMIO IOANNINON (UOI)

Address: PANEPISTEMIOYPOLE PANEPISTEMIO IOANNINON, IOANNINA 45110, Greece

Title of the department (s): Department of Materials Science and Engineering, Unit of Medical Technology and Intelligent Information Systems

Principal investigator: Prof. Dimitrios I. Fotiadis, e-mail: fotiadis@uoi.gr

Research team: Eleni Georga (egeorga@uoi.gr), Marina Georgoula (mgeorgoula@uoi.gr), Vasilis Aidonis (vaidonis@uoi.gr), Orestis Papagiannopoulos (orepap@uoi.gr), Daphne Katsarou (d.katsarou@uoi.gr), Konstantinos Mavrokotas (k.mavrokotas@uoi.gr)

DPO: Stavroula Stathara (dpo@uoi.gr)

Reason for data access: data curation, data harmonization, data pre-processing, machine learning and deep learning predictive modeling, data anonymization / pseudonymization.

Research partner No 4: Computer Solutions Cyprus LTD (CSCY)

Address: Diagorou 4, KERMIA BUILDING, 8th fl. 802, 1097, Nicosia, Cyprus

Title of the department (s): R&D

Principal investigator: Dr Stavros Pitoglou (s.pitoglou@csl.gr)

Research team: Dr Thelma Androutsou (t.androutsou@csl.gr), Ioannis-Andreas Filippas (i.filippas@csl.gr)

DPO: Athanasios Spetsarias (a.spetsarias@csl.gr)

© 2023-2025 BIO-STREAMS

Reason for data access: data API development, data curation, data harmonization, data pre-processing, machine learning and deep learning predictive modeling, data anonymization / pseudonymization.

Research partner No 5: Novelcore (NVCR)

Address: Mavromichali 104, Athina 114 72

Title of the department (s): Novelcore

Principal investigator: Georgios Domalis (domalis@novelcore.eu)

Research team: Ioannis Livieris (livieris@novelcore.eu), Nikolaos Alimpertis (alimpertis@novelcore.eu)

DPO: Dimitrios Charalampakis (charalampakis@novelcore.eu),

Reason for data access: data pre-processing, data enrichment, synthetic data generation, knowledge graph creation, data evaluation.

Research partner No 6: Ainigma technologies (AINIGMA)

Address: Kapeldreef 60 - Leuven 3001, Belgium

Title of the department (s): RnD

Principal investigator: Athanasios Kakasis (kakasisathan@ainigma.tech)

Research team: Christos Chatzichristos (cchatzic@ainigma.tech), Ali Saad (a.saad@ainigma.tech), Marianna Panagiotidou (m.panagiotidou@ainigma.tech)

DPO: Athanasios Kakasis (kakasisathan@ainigma.tech)

Reason for data access: data pre-processing, machine learning and deep learning predictive modeling

Research partner No 7: HAROKOPIO UNIVERSITY OF ATHENS (HUA)

Address: El. Venizelou 70, Kallithea, 176 71 Athens, Greece

Title of the department(s): Department of Informatics and Telematics

Principal investigator: Associate Prof. George Dimitrakopolos, e-mail: gdimitra@hua.gr

Research team: George Dimitrakopoulos (gdimitra@hua.gr), Ilias Panagiotopoulos (ipanagio@hua.gr), Christina-Athanasia Alexandropoulou (calexand@hua.gr), Yioula Lekka (ylekka@hua.gr), Despoina Mitsiogianni (it219146@hua.gr)

Reason for data access: data curation, data harmonization, data pre-processing, machine learning and deep learning predictive modeling, data anonymization / pseudonymization.

Research partner No 8: Innovation to Grow (i2G)

Address: Via A. Appiani 12, 20121 Milano, Italy

© 2023-2025 BIO-STREAMS

Title of the department (s): Innovation Dept.

Principal investigator: Umberto Restelli (u.restelli@i2grow.it)

Research team: Sofia Silvola (s.silvola@i2grow.it),

DPO: Matteo Colombo (m.colombo@i2grow.it)

Reason for data access: assessment of resources used and cost data. Analysis of available data for health economic indicators implementation

Research partner No 9: UKEMED GLOBAL Ltd (UKEMED)

Address: 121, Prodromou Street Offices 713-715 2064, Nicosia, Cyprus

Title of the department (s): Innovation and Research

Principal investigator: Iliana Korma (i.korma@gmail.com)

Research team: Zenia Koti (zeniakoti@ukemed.com), Chara Kapsala (chara.kapsala@ukemedglobal.com)

DPO: Takis Kotis (takiskotis@ukemeglobal.com)

Reason for data access: assessment of resources used and cost data. Analysis of available data for health economic indicators implementation, public dissemination.

Research partner No 10: TECREANDO B.V. (TCR)

Address: Herengracht 575-577, 1017 CD, Amsterdam, the Netherlands

Title of the department (s): N/A

Principal investigator: Sotiris Pavlopoulos (pavlopoulos@tecreando.com)

Research team: Ioannis Vezakis (vezakis@tecreando.com)

DPO: Panagiotis Tsatsoulis (tsatsoulis@tecreando.com)

Reason for data access: Data harmonisation, data analysis, machine learning and predictive modeling

Research partner No 11: EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLGISTON-EMP (ICCS)

Address: PATISION 42, ATHINA 106 82, Greece

Title of the department (s): Biomedical Engineering Laboratory (BEL) of the School of Electrical and Computer Engineering (ECE) of the National Technical University of Athens (NTUA)

Principal investigator: Prof. D. Koutsouris (dkoutsou@biomed.ntua.gr)

Research team: E. Vellidou (ebel@biomed.ntua.gr), I. Kakkos (ikakkos@biomed.ntua.gr), I. Kouris (ikouris@biomed.ntua.gr), A. Anastasiou (aanastasiou@biomed.ntua.gr), K. Bromis (konbromis@biomed.ntua.gr), V. Apostolakos (vapostolakos@biomed.ntua.gr)

© 2023-2025 BIO-STREAMS

DPO: Dimitris Kalogeras (dkalog@gmail.com)

Reason for data access: Data collection, data curation, data harmonization, data pre-processing, machine learning and deep learning predictive modeling, data anonymization / pseudonymization.

© 2023-2025 BIO-STREAMS

7 Appendix B: Informed Consent

7.1 INFORMED CONSENT FOR CLINICAL EXPERT PANEL

7.1.1 Study Title:

European Childhood Obesity Platform: Standardized Data Harmonization and AI Model Validation

7.1.2 Principal Investigator(s):

[Name of Principal Investigator]
[Institution]
[Contact Information]

7.1.3 Co-Investigators:

[Names of Co-Investigators and their institutions]

7.1.4 INVITATION TO PARTICIPATE

Dear Expert,

You are being invited to participate as a member of the Clinical Expert Panel in the EU BIO-STREAMS project. One of the aims of the project is to develop and validate ML/AI models for predicting childhood obesity risk factors and outcomes using harmonized data from seven European clinical sites.

Before you decide whether to participate, please take time to read the following information carefully. Feel free to discuss it with colleagues or ask us if there is anything that is not clear or if you would like more information.

7.1.5 PROJECT OVERVIEW

The BIO-STREAMS project is creating a standardized virtual biobank of childhood obesity data from multiple European clinical sites. This data will be used to develop ML/AI models that can identify risk factors and predict outcomes for children aged 5-18 years. As part of the validation process, we require experienced clinicians to evaluate the practical utility and clinical relevance of these models.

7.1.6 YOUR ROLE AS AN EXPERT

As a member of the Clinical Expert Panel, you will focus on validating the practical utility of the prediction models. Your specific responsibilities will include:

1. Performing independent clinical assessments on anonymized test cases without knowledge of AI predictions

© 2023-2025 BIO-STREAMS

2. Evaluating whether risk factors identified by AI models align with established obesity pathways
3. Assessing the clinical meaningfulness of AI-generated explanations
4. Validating the real-world applicability of model outputs in patient care
5. Providing your expert opinion on how these models could impact treatment decisions
6. Completing standardized evaluation forms documenting your assessments

Your participation will involve two rounds of blind testing:

- **Round 1**: October-December 2025 (approximately 20-30 hours of work)
- **Round 2**: August-October 2026 (approximately 20-30 hours of work)

You will be provided with standardized evaluation forms and access to a secure platform for reviewing cases and recording your assessments.

7.1.7 ELIGIBILITY AND SELECTION

You have been invited to participate because you are:

- A practicing clinician with expertise in pediatric medicine, endocrinology, or related fields.
- Experienced in childhood obesity management and treatment.
- Either currently affiliated with one of the seven participating clinical sites (internal expert) OR completely independent from the project (external expert).

7.1.8 CONFIDENTIALITY AND DATA PROTECTION

All patient data you will review has been pseudonymized in compliance with GDPR and other applicable regulations. Your assessments and feedback will be identified only by a unique participant ID, not by your name. Your personal information will be stored separately from your evaluations and will only be accessible to authorized members of the research team.

The project follows strict data protection protocols. You will be required to:

- Not download or share any patient data
- Maintain confidentiality of all information viewed during the study
- Complete all evaluations in private settings

7.1.9 BENEFITS AND RISKS

Potential Benefits:

- Contributing to advancements in childhood obesity prediction and management
- Exposure to innovative AI approaches that may enhance clinical practice
- Professional development through participation in a major European research initiative
- Acknowledgment in project publications (with your consent)

© 2023-2025 BIO-STREAMS

- Access to research findings before wider dissemination

Potential Risks:

- The time commitment required for thorough evaluation
- Potential professional disagreement with AI assessments that may cause minor frustration

7.1.10 VOLUNTARY PARTICIPATION

Your participation in this study is entirely voluntary. You are free to withdraw at any time without giving any reason and without any negative consequences. If you choose to withdraw, you may decide whether data collected up to that point can be used in the study.

7.1.11 PUBLICATION AND ACKNOWLEDGMENT

Results from this study will be published in scientific journals and presented at conferences. Your contribution will be acknowledged in publications if you consent to being named. If you prefer, your contribution can remain anonymous with only your role (e.g., "clinical expert") being mentioned.

7.1.12 ETHICS APPROVAL

This study has been reviewed and approved by [name of Ethics Committee/IRB] (Reference number: [reference]).

7.1.13 CONSENT DECLARATION

By signing below, I confirm that:

- I have read and understood the information provided about the study.
- I have had the opportunity to ask questions and have received satisfactory answers.
- I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.
- I agree to participate as a Clinical Expert Panel member in both rounds of blind testing.
- I understand how my data will be collected, used, and stored for this research.
- I agree to maintain the confidentiality of all patient information I review.
- I will provide my clinical assessment independently and according to my professional judgment.

Please indicate your preferences for acknowledgment in publications:

- [ ] I consent to being named as a clinical expert contributor in publications
- [ ] I prefer to remain anonymous in publications (role only mentioned)

Please indicate your status:

- [ ] Internal Expert (affiliated with one of the seven participating clinical sites)

© 2023-2025 BIO-STREAMS

- [ ] External Expert (independent from the project)

Participant:

Name: _____

Professional Title: _____

Institution: _____

Clinical Specialty: _____

Years of Experience in Obesity Management: _____

Signature: _____ Date: _____

Researcher obtaining consent:

Name: _____

Signature: _____ Date: _____

This document is to be signed in duplicate, with one copy retained by the participant and one by the research team.

8 CONTACT INFORMATION

If you have any questions or concerns about the study, please contact:

Study Coordinator:
[Name]
[Email]
[Phone Number]

Clinical Panel Coordinator:
[Name]
[Email]
[Phone Number]

Data Protection Officer:
[Name]
[Email]
[Phone Number]

Ethics Committee:
[Contact Information]

© 2023-2025 BIO-STREAMS

8.1 INFORMED CONSENT FOR RESEARCH EXPERT PANEL

8.1.1 Study Title:

European Childhood Obesity Platform: Standardized Data Harmonization and AI Model Validation

8.1.2 Principal Investigator(s):

[Name of Principal Investigator]
[Institution]
[Contact Information]

8.1.3 Co-Investigators:

[Names of Co-Investigators and their institutions]

8.1.4 INVITATION TO PARTICIPATE

Dear Researcher,

You are being invited to participate as a member of the Research Expert Panel in the EU Project BIO-STREAMS. One of the aims of the project is to develop and validate ML/AI models for predicting childhood obesity risk factors and outcomes using harmonized data from seven European clinical sites.

Before you decide whether to participate, please take time to read the following information carefully. Feel free to discuss it with colleagues or ask us if there is anything that is not clear or if you would like more information.

8.1.5 PROJECT OVERVIEW

The EU Childhood Obesity Research Project is creating a standardized virtual biobank of childhood obesity data from multiple European clinical sites. This data will be used to develop AI models that can identify risk factors and predict outcomes for children aged 5-18 years. As part of the validation process, we require experienced researchers to evaluate the methodological rigor and technical validity of these models.

8.1.6 YOUR ROLE AS A RESEARCH EXPERT

As a member of the Research Expert Panel, you will focus on carrying out scientific validation of the AI models. Your specific responsibilities will include:

1. Evaluating the technical robustness of the ML/AI models
2. Analyzing the statistical validity of predictions against established metrics
3. Assessing explainability frameworks (SHAP, LIME) from a scientific perspective

© 2023-2025 BIO-STREAMS

4. Validating methodological approaches in model development
5. Reviewing model architectures, training protocols, and anonymized prediction outputs
6. Completing standardized assessment frameworks documenting your technical evaluations

Your participation will involve two rounds of validation:

- **Round 1**: October-December 2025 (approximately 15-20 hours of work)
- **Round 2**: August-October 2026 (approximately 15-20 hours of work)

You will be provided with model architecture documentation, training protocols, evaluation metrics, and access to a secure platform for reviewing model performance and recording your assessments.

8.1.7 ELIGIBILITY AND SELECTION

You have been invited to participate because you are:

- A researcher with expertise in machine learning, artificial intelligence, biostatistics, or related fields.
- Experienced in predictive modeling, particularly in healthcare applications.
- Familiar with model validation methodologies and evaluation frameworks.
- Either currently affiliated with one of the seven participating research institutions (internal expert) OR completely independent from the project (external expert).

8.1.8 CONFIDENTIALITY AND DATA PROTECTION

All data you will review has been pseudonymized in compliance with GDPR and other applicable regulations. Your assessments and feedback will be identified only by a unique participant ID, not by your name. Your personal information will be stored separately from your evaluations and will only be accessible to authorized members of the research team.

The project follows strict data protection protocols. You will be required to:

- Access data and models only through the secure project platform
- Not download or share any patient data or proprietary model information
- Maintain confidentiality of all information reviewed during the study
- Complete all evaluations in secure settings

8.1.9 BENEFITS AND RISKS

Potential Benefits:

- Contributing to advancements in ML/AI applications for healthcare
- Exposure to innovative methodological approaches in predictive modeling
- Professional development through participation in a major European research initiative

© 2023-2025 BIO-STREAMS

- Acknowledgment in project publications (with your consent)
- Access to research findings and methodologies before wider dissemination

Potential Risks:

- The time commitment required for thorough technical evaluation
- Potential intellectual disagreement with methodological approaches

8.1.10 VOLUNTARY PARTICIPATION

Your participation in this study is entirely voluntary. You are free to withdraw at any time without giving any reason and without any negative consequences. If you choose to withdraw, you may decide whether data collected up to that point can be used in the study.

8.1.11 PUBLICATION AND ACKNOWLEDGMENT

Results from this study will be published in scientific journals and presented at conferences. Your contribution will be acknowledged in publications if you consent to being named. If you prefer, your contribution can remain anonymous with only your role (e.g., "research expert") being mentioned.

8.1.12 INTELLECTUAL PROPERTY

Your participation in this validation process does not grant you intellectual property rights to the models or methodologies developed in the project. However, your feedback may contribute to improvements in the models, and your contribution will be appropriately acknowledged in resulting publications.

8.1.13 ETHICS APPROVAL

This study has been reviewed and approved by [name of Ethics Committee/IRB] (Reference number: [reference]).

8.1.14 CONSENT DECLARATION

By signing below, I confirm that:

- I have read and understood the information provided about the study.
- I have had the opportunity to ask questions and have received satisfactory answers.
- I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.
- I agree to participate as a Research Expert Panel member in both rounds of validation.
- I understand how my data will be collected, used, and stored for this research.
- I agree to maintain the confidentiality of all information I review.
- I will provide my scientific assessment independently and according to established methodological standards.

© 2023-2025 BIO-STREAMS

Please indicate your preferences for acknowledgment in publications:

- [ ] I consent to being named as a research expert contributor in publications
- [ ] I prefer to remain anonymous in publications (role only mentioned)

Please indicate your status:

- [ ] Internal Expert (affiliated with one of the seven participating research institutions)
- [ ] External Expert (independent from the project)

Participant:

Name: _____

Professional Title: _____

Institution: _____

Research Specialty: _____

Years of Experience in ML/AI or Biostatistics: _____

Signature: _____ Date: _____

Researcher obtaining consent:

Name: _____

Signature: _____ Date: _____

This document is to be signed in duplicate, with one copy retained by the participant and one by the research team.

8.1.15 CONTACT INFORMATION

If you have any questions or concerns about the study, please contact:

Study Coordinator:
[Name]
[Email]
[Phone Number]

Research Panel Coordinator:
[Name]
[Email]
[Phone Number]

Data Protection Officer:
[Name]
[Email]
[Phone Number]

Ethics Committee:
[Contact Information]

© 2023-2025 BIO-STREAMS

9 Appendix C: Data Transfer/Processing Agreements

Pursuant to article 28, par. 3 of GDPR, for the purposes of the **BIO-STREAMS** Project, a partner in its quality of **Data Controller** and partners in their quality of **Data Processor** that process Personal Data of the **Data Controller** execute the following:

Data Processing Agreement

By and between:

Data Controller: XXX, established in **XXX** (hereinafter the "**Data Controller**"),

and

Data Processors

Data Processors

Name	**Full address**
Faculty of Electrical Engineering and Computer Science, University of Maribor (UM)	Koroška cesta 46, 2000 Maribor, Slovenia
PANEPISTIMIO IOANNINON (UOI)	PANEPISTEMIOYPOLE PANEPISTEMIO IOANNINON, IOANNINA 45110, Greece
Computer Solutions Cyprus LTD (CSCY)	Diagorou 4, KERMIA BUILDING, 8th fl. 802, 1097, Nicosia, Cyprus
Novelcore (NVCR)	Parodos Theofrastou 140, 26443, Patras, Greece
AINIGMA Technologies (AINIGMA)	Innovation & Incubation Center KU Leuven, Kapeldreef 60, Leuven 3000, Belgium
HAROKOPIO UNIVERSITY OF ATHENS (HUA)	70, El. Venizelou street, Kallithea, Greece, 17671
Innovation to Grow (I2G)	Via A. Appiani 12, 20121 Milano, Italy
UKEMED GLOBAL Ltd (UKEMED)	121, Prodromou Street, 2064, Nicosia, Cyprus

© 2023-2025 BIO-STREAMS

TECREANDO B.V. (TCR)	Herengracht 575-577, 1017 CD, Amsterdam, Netherlands
EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLOGISTON-EMP (ICCS)	42 Patision Str, 10682 Athens, Greece

(hereinafter the "Technical Partner"/ "Technical Partners" or the "**Data Processor**" "Data Processors", and together with the Data Controller, the "**Parties**")

WHEREAS:

A. XXX acts as a **Data Controller.**
B. For the purpose of the joint research project activities, the Technical Partners, acting as **Data Processors** will process Controller's Personal Data on behalf of the **Data Controllers**.
C. The Parties seek to implement this **Data Processing Agreement** ("**DPA**") that complies with the requirements of the current legal framework in relation to data processing and with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (hereinafter, "**GDPR"**).
D. By the execution of the present DPA, the Parties wish to lay down their rights and obligations.
E. The Data Controller is also subject to the rights and obligations as a separate "data controller" set forth under the GDPR in relation to the processing of personal data of its patients for purposes other than conducting the BIO-STREAMS Project. In particular, the Data Controller remains data controller of the data contained in its patients' medical records for the purposes of providing medical care to its patients and for academic research purposes.

IT IS AGREED AS FOLLOWS:

1. Definitions and Interpretation

Unless otherwise defined, capitalised terms and expressions used in this Agreement shall have the following meaning:

"**Agreement"** and "DPA" means this Data Processing Agreement and all Schedules;

"**Controller's Personal Data**" means any Personal Data Processed by the Data **Processor** on behalf of the Data Controller pursuant to or in connection with the processing of the personal patient data retrieved from the Patients records and Collected data from the Clinical study described in Clinical study protocol;

© 2023-2025 BIO-STREAMS

“Personal Data” means any information relating to an identified or identifiable natural person ("Data Subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

"Personal Data Breach" means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, Personal Data;

"Processing" means any operation or set of operations which is performed on Personal Data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

"**Data Protection Laws**" means EU Data Protection Laws and, to the extent applicable, the data protection or privacy laws of any other country applicable to the specific case;

"**EEA**" means the European Economic Area;

"**GDPR**" means Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, or EU General Data Protection Regulation;

"**BIO-STREAMS**" means joint EU funded research project of which the Parties are members of the consortium;

"**Data Transfer**" means:

- a transfer of Data **Controller**’s Personal Data to Data **Processor**; or
- an onward transfer of Data Controller’s Personal Data from a Data **Processor** to a Sub**-processor**, or between two Data **Processors**,
- in each case, where such transfer would be prohibited by Data Protection Laws (or by the terms of data transfer agreements put in place to address the data transfer restrictions of Data Protection Laws);

"Processor" means a natural or legal person, public authority, agency or other body which processes Personal Data on behalf of the Controller;

"**Sub-processor**" means any person appointed by or on behalf of **Processor** to process Personal Data on behalf of the Data Controller in connection with the Agreement.

The terms, "**Commission**", "**Controller**", "**Data Subject**", "**Member State**", and "**Supervisory Authority**" shall have the same meaning as in the GDPR, and their cognate terms shall be construed accordingly.

2. Processing of Personal Data

Processor shall:

a. comply with all applicable Data Protection Laws in the Processing of **Controller’s** Personal Data; and
b. not process **Controller’s** Personal Data other than on the relevant **Controller**’s documented instructions

© 2023-2025 BIO-STREAMS

The **Controller** instructs the Processor to process the **Controller**'s Personal Data.

Processor shall:

a. comply with all applicable Data Protection Laws in the Processing of **Controller's** Personal Data; and
b. not process **Controller's** Personal Data other than on the relevant **Controller**'s documented instructions

The Processors are instructed to process the Personal Data for the term of this Agreement and only for the purposes of providing the data processing tasks set out in Annex 1. As part of the Bio-Streams, **Processors** may not process or use Personal Data in another way than provided in the instructions, including with regard to transfers of Personal Data to a third country or an international organisation, unless **Processor** is required to do so according to Union or Member State law, in particularly GDPR, Article 46. In that case, **Processor** shall inform the **Controller** in writing of that legal requirement before processing, unless that law prohibits such information on important grounds of public interest.

Processor shall at all times maintain a record of processing of Personal Data in accordance with applicable law and, if **Processor** considers an instruction from **Controller** to be in violation of the applicable law, **Processor** shall promptly inform **Controller** in writing.

3. Processor Personnel - Confidentiality

Processor shall take reasonable steps to ensure the reliability of any employee, who may have access to the **Controller**'s Personal Data, ensuring in each case that access is strictly limited to those individuals who need to know/access the relevant **Controller**'s Personal Data, as strictly necessary, and to comply with Applicable Laws in the context of that individual's duties to the **Processor**, ensuring that all such individuals are subject to confidentiality undertakings or professional or statutory obligations of confidentiality.

4. Security

Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of Processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, **Processor** shall in relation to the **Controller's** Personal Data implement appropriate technical and organisational measures to ensure a level of security appropriate to that risk of processing.

Processor shall implement appropriate technical and organisational measures to prevent that the Personal Data processed is:

- accidentally or unlawfully destroyed, lost or altered,
- disclosed or made available without authorisation, or
- otherwise processed in violation of applicable law,

In assessing the appropriate level of security, the **Processor** shall take account of the risks that are presented by Processing, in particular from a Personal Data Breach.

5. Sub-processing

Processor shall not appoint (or disclose any **Controller's** Personal Data) to any Sub-processor unless required or authorised by the **Controller.**

Processor may only engage a Sub-processor, with prior specific or general written consent from the Controller. Processor undertakes to inform Controller of any intended changes

© 2023-2025 BIO-STREAMS

concerning the addition or replacement of a sub-processor by providing a reasonable prior written notice to Controller. Controller may reasonably and in a duly substantiated manner object to the use of a Sub-processor. Processor must inform the Controller in writing of the discontinued use of a Sub-processor.

Prior to the engagement of a Sub-processor, Processor shall conclude a written agreement with the Sub-processor, in which at least the same data protection obligations as set out in this Data Processing Agreement shall be imposed on the Sub-processor, including obligations to implement appropriate technical and organisational measures and to ensure that the transfer of Personal Data is done in such a manner that the processing will meet the requirements of the Applicable Law.

Controller has the right to receive a copy of the relevant provisions of Processor's agreement with the Sub-processor related to data protection obligations. Processor shall remain fully liable to Controller for the performance of the Sub-processors' obligations under this Data Processing Agreement. The fact that the Controller has given consent to the Data Processor's use of a Sub-processor is without prejudice for the Data Processor's duty to comply with this Data Processing Agreement.

6. Duty of Assistance and Data Subject Rights

Processor will assist the Controller in ensuring compliance with its obligations as a Controller.

The Parties undertake to cooperate with the competent data protection authorities, in particular in the event of a request for information which may be addressed to them or in the event of an inspection.

Taking into account the nature of the Processing, **Processor** shall assist the **Controller** by implementing appropriate technical and organisational measures, insofar as this is possible, for the fulfilment of the **Controller** obligations, as reasonably understood by **Controller**, to respond to requests to exercise Data Subject rights under the Data Protection Laws. **Processor** shall:

a. Promptly notify **Controller** if it receives a request from a Data Subject under any Data Protection Law in respect of **Controller's** Personal Data; and
b. Ensure that it does not respond to that request except on the documented instructions of **Controller** or as required by Applicable Laws to which the **Processor** is subject, in which case **Processor** shall to the extent permitted by Applicable Laws inform **Controller** of that legal requirement before the **Processor** responds to the request.

7. Personal Data Breach

Processor shall notify **Controller** without undue delay, and at the latest within twenty-four (24) hours, upon **Processor** becoming aware of a Personal Data Breach affecting **Controller**'s Personal Data, providing **Controller** with sufficient information to allow the **Controller** to meet any obligations to report or inform Data Subjects of the Personal Data Breach under the Data Protection Laws.

Processor will provide Controller with the following information regarding the Personal Data Breach at the time of notification of the incident, or if this is not reasonably possible as soon as possible after the notification of the Personal Data Breach, including but not limited to:

- The nature of the Personal Data Breach, stating the categories and (by approximation) the number of Data Subjects concerned, and stating the categories and (by approximation) the number of the personal data registers affected (datasets);
- The likely consequences of the Personal Data Breach;

© 2023-2025 BIO-STREAMS

- Steps taken to address the Personal Data Breach, including (where appropriate) measures to mitigate any possible adverse effects of such breach.

Processor shall cooperate with the **Controller** and take reasonable steps as directed by the **Controller** to assist in the investigation, mitigation and remediation of each such Personal Data Breach.

8. Data Protection Impact Assessment and Prior Consultation

Processor shall provide reasonable assistance to the **Controller** with any data protection impact assessments, and prior consultations with Supervising Authorities or other competent data privacy authorities, which **Controller** reasonably considers to be required by article 35 or 36 of the GDPR or equivalent provisions of any other Data Protection Law, in each case solely in relation to Processing of **Controller's** Personal Data by, and taking into account the nature of the Processing and information available to, the **Processor**.

9. Deletion or return of Controller's Personal Data

Processor shall promptly and in any event within 10 business days of the date of cessation of BIO-STREAMS project involving the Processing of **Controller's** Personal Data (the "**Cessation Date**"), delete and procure the deletion of all copies of those **Controller** Personal Data. **Processor** shall provide written certification to **Controller** that it has fully complied with this section 9 within 10 business days of the Cessation Date.

10. Audit Rights

Processor shall make available to the **Controller** on request information necessary to demonstrate compliance with this Agreement, and shall allow for and contribute to audits, including inspections, by the **Controller** or an auditor mandated by the **Controller** in relation to the Processing of the **Controller** Personal Data.

Information and audit rights of the **Controller** only arise under section 10 to the extent that the Agreement does not otherwise give them information and audit rights meeting the relevant requirements of Data Protection Law.

11. Data Transfer

The **Processor** may not transfer or authorise the transfer of Data to countries outside the EU and/or the European Economic Area (EEA) without the prior written consent of the **Controller**. If personal data processed under this Agreement is transferred from a country within the European Economic Area to a country outside the European Economic Area, the Parties shall ensure that the personal data are adequately protected. To achieve this, the Parties shall, unless agreed otherwise, rely on EU approved standard contractual clauses for the transfer of personal data.

12. Responsibility

The Parties are each responsible for their own actions.

Processor shall be liable for and indemnify Controller against all damages and claims from third parties, including the Person concerned, resulting from a breach by Processor of this Agreement and of the obligations specifically addressed to Processor in the GDPR.

The degree of liability for breaches arising from this agreement is limited to the agreements concluded in the main agreement.

© 2023-2025 BIO-STREAMS

13. General Terms

Confidentiality. Each Party must keep this Agreement and information it receives about the other Party and its business in connection with this Agreement ("**Confidential Information**") confidential and must not use or disclose that Confidential Information without the prior written consent of the other Party except to the extent that:

a. disclosure is required by law;
b. the relevant information is already in the public domain.

Notices. All notices and communications given under this Agreement must be in writing and will be delivered personally, sent by post or sent by email to the address or email address set out in the heading of this Agreement at such other address as notified from time to time by the Parties changing address.

Governing Laws and Jurisdiction. This Agreement is governed by the laws of XXX.

Any dispute arising in connection with this Agreement, which the Parties will not be able to resolve amicably, will be submitted to the exclusive jurisdiction of the courts of XXX.

The Agreement shall become effective separately between each Processor and the Controller on the date where both the Controller and the Processor have provided valid signature ("Effective Date"). The Controller shall provide access to Personal Dara to each Processor upon their respective Effective Date.

© 2023-2025 BIO-STREAMS

Signatures

Role	Organisation	Authorised Signatory	Signature and Stamp
Data Controller	XXX	XXX	
Data Processor	University of Maribor	Prof. Dr. Zdravko Kačič,rector of UM	
Data Processor	PANEPISTIMIO IOANNINON (UOI)	Prof. Anna Batistatou, Rector of UOI	
Data Processor	Computer Solutions Cyprus LTD (CSCY)	Mrs Vasiliki Paziana	
Data Processor	Novelcore (NVCR)	Mr Dimitrios Tsakalidis	
Data Processor	AINIGMA Technologies (AINIGMA)	Athanasios Kakasis, CEO	
Data Processor	HAROKOPIO UNIVERSITY OF ATHENS (HUA)	Prof. George Dimitrakopoulos	
Data Processor	Innovation to Grow (j2G)	Mr. Matteo Colombo, Legal representative	
Data Processor	UKEMED GLOBAL Ltd (UKEMED)	Mr Takis Kotis Legal representative	
Data Processor	TECREANDO B.V. (TCR)	Mr Panagiotis Tsatsoulis	
Data Processor	EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLGISTON-EMP (ICCS)	Dr. Dimitrios Kalogeras	

© 2023-2025 BIO-STREAMS

9.1 Data Processor: University of Maribor

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Study Oversight and Management, Risk and Adverse effects Management, Data Analysis and Interpretation, Reporting and Publication

☐ Other: (specify)

Protocol number: .....

SERVICES DESCRIPTION

Study Oversight, Risk and Adverse effects Management, Data analysis and interpretation and scientific dissemination.

INTENDED PROCESSING ACTIVITIES

☐ Collection
☐ Registration
☐ Organisation
☐ Structuring
☐ Conservation
☐ Adaptation or modification
☑ Extraction
☐ Consultation
☑ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☐ Other (specify):

© 2023-2025 BIO-STREAMS

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

- ☑ Data concerning health:
- ☐ Data revealing trade union membership
- ☑ Data revealing racial or ethnic origin
- ☐ Data revealing political opinions
- ☐ Data revealing religious or philosophical beliefs
- ☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

As a Clinical Coordinators we will aquire and process data elements as specified under BIO-STREAMS common semantic data model including: healthcare indices, demographics, behavioral data, medical data, and family data.

PLACE OF THE PROCESSING

Infrastructure of the University of Maribor, Slovenia. The managed computers are connected to the local secured network and hosted and managed by the University of Maribor.

THIRD PARTIES

N/A

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Izidor Mlakar (izidor.mlakar@um.si)	Izidor Mlakar (izidor.mlakar@um.si)

© 2023-2025 BIO-STREAMS

	DPO: assoc. prof. dr. Miha Dvojmoč (dpo@um.si)	DPO: assoc. prof. dr. Miha Dvojmoč (dpo@um.si)

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Antivirus and Anti-malware software
- A central storage system with regular archiving and backups
- All point of entries into internal network are protected by central policy and service
- Connection traceability (log file)
- User authentication and access to information resources control

Organisational measures could be one of the following

- Physical access control: limited access to the infrastructure only within physical network (or using VPN with the digital identity)
- Logging (logging processes implemented on the institutional level as defined in Universty Of Maribor's Information Security Policy)
- Information security policy: Universty Of Maribor's Information Security Policy
- Access management: Digital Identity
- Directory of those processing personal data and limiting access to research data (including personal data) only to authorized researchers with legal baseline (i.e. specifically mentioned in the protocol)
- Yearly evaluation and testing of the ICT measures in place, Compatibility and Security Scans for Web servers every 3months, for Internal production servers every 6 months and Other internal servers and equipment once a year.
- Disaster recovery plan: Universty Of Maribor's Information Security Policy
- Procedure for managing security incidents, data breaches and business continuity Universty Of Maribor's Information Security Policy, Protection of personal data - UM.si

9.2 Data Processor: PANEPISTIMIO IOANNINON (UOI)

Annex I – Instructions and Processing Details

© 2023-2025 BIO-STREAMS

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Data curation, data harmonisation, data pseudonymization, Artificial Intelligence/Machine Learning (AI/ML)-based analytics.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☐ Prospective research

Protocol number: ……

SERVICES DESCRIPTION

Data curation (i.e., assessment of data quality properties inclusive of representation bias), data harmonisation (i.e., representation of data based on the BIO-STREAMS common data model and data ontology), data pseudonymization, development and validation of AI/ML-based risk prediction models of adverse metabolic outcomes attributed to obesity and overweight, and temporary data storage on the PRECIOUS HPC Infrastructure at the University of Ioannina (Research infrastructure for big medical data analytics towards precision medicine, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund)).

INTENDED PROCESSING ACTIVITIES

☐ Collection
☐ Registration
☑ Organisation
☑ Storage
☑ Structuring
☐ Conservation
☑ Adaptation or modification
☑ Extraction
☑ Consultation
☑ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

© 2023-2025 BIO-STREAMS

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

☑ Data concerning health:

☐ Data revealing trade union membership

☑ Data revealing racial or ethnic origin

☐ Data revealing political opinions

☐ Data revealing religious or philosophical beliefs

☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

BIO-STREAMS retrospectively (Study 1) collected data including: healthcare indices, demographics, behavioural data, medical data, and family data.

© 2023-2025 BIO-STREAMS

PLACE OF THE PROCESSING

Data processing will be conducted on the PRECIOUS HPC Infrastructure (Research infrastructure for big medical data analytics towards precision medicine, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund)). PRECIOUS is a cloud-based supercomputing infrastructure of the Unit of Medical Technology and Intelligent Information Systems at the University of Ioannina which can provide services to healthcare actors on local, national and international level. PRECIOUS provides exceptional performance assistance for challenging scientific applications in the healthcare industry. The infrastructure's primary capability is its ability to support large data management, cloud storage, cloud computing apps, and high-performance computing computations all at once.

THIRD PARTIES

N/A

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Prof. Anna BATISTATOU, Rector of UOI DPO: Mrs. Stavroula Stathara, dpo@uoi.gr, +30 265 100 7321	Prof. Anna BATISTATOU, Rector of UOI DPO: Mrs. Stavroula Stathara, dpo@uoi.gr, +30 265 100 7321

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities
- Information security policy
- Access management: 2-factor authentication for example (something you have + password + login)
- Directory of those processing personal data
- Process for regularly testing, assessing and evaluating
- Privileged identity management: minimise access to personal data
- Logging
- Disaster recovery plan
- Procedure for managing security incidents and data breaches
- Paper document security

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

The PRECIOUS HPC Infrastructure supports all the aforementioned security measures both at the infrastructure level and at the VM level. PRECIOUS with the VxRail technology provides a wide range of security capabilities to protect the data infrastructure from cyberattacks or ensure that the data is protected if a successful breach occurs. There are many layers of protection, from the security features embedded in the PowerEdge hardware, to the virtualization software developed by VMware in the Application Layer, to the VxRail HCI software which serves as a secure link between these layers. Together, these work to protect critical components, such as the BiOS,

© 2023-2025 BIO-STREAMS

Firmware, and the data stored in VSAN. Some of the key features (not an exhaustive set) are: User Authentication and Authorization, Secure Root of Trust, vSAN encryption, Signed LCM update bundles, STIG Hardening. In particular “Pseudonymization and encryption of personal data and encryption of communications in case of transfer” and “Access management: 2-factor authentication for example (something you have + password + login)” are supported in VMs but also at the application level.

9.3 Data Processor: Computer Solutions Cyprus LTD (CSCY)

Annex I – Instructions and Processing Details

9.3.1.1.1.1 NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: 1) Development and deployment of the Biobank Node Bundles (BNBs): Each BNB is a Data Hub, including the necessary computer resources for deployment and interconnection within the Bio-Streams Hub Network. This network ensures that the BIO-STREAMS-Biobank is based on a decentralized federated data management system. 2) Participation in Data Harmonization and Data Curation processes, assessing the quality of data and creating a common data model and data ontology for BIO-STREAMS.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: ……

9.3.1.1.1.2

9.3.1.1.1.3 SERVICES DESCRIPTION

The Node Bundles accept data that is uploaded from authorized users of the clinical partners, and after data processing (harmonization, anonymization) they store them locally and make them available via API endpoints in the Information Management System (IMS).

9.3.1.1.1.4 INTENDED PROCESSING ACTIVITIES

☑ Collection
☐ Registration
☑ Organisation
☑ Structuring
☑ Conservation
☑ Adaptation or modification

© 2023-2025 BIO-STREAMS

☑ Extraction
☐ Consultation
☐ Use
☑ Distribution or any other form of making available
☑ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

☑ Data concerning health:

☐ Data revealing trade union membership

☑ Data revealing racial or ethnic origin

☐ Data revealing political opinions

☐ Data revealing religious or philosophical beliefs

☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

BIO-STREAMS retrospectively (Study 1) and prospectively (Study 2 & Study 3) collected data including: healthcare indices, demographics, behavioural data, medical data, and family data.

© 2023-2025 BIO-STREAMS

PLACE OF THE PROCESSING (Greece)

The Biobank Node Bundles reside inside the clinical sites involved in the project. All data are stored, processed and accessed there.

THIRD PARTIES

N/A

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Mr Athanasios Spetsarias a.spetsarias@csl.gr	Mr Athanasios Spetsarias a.spetsarias@csl.gr

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

T.O.M.P. (Technical and Organizational Protection Measures)

The following refer to measures and procedures applied by Computer Solutions (CSCY) and its staff to comply with the General Data Protection Regulation (GDPR), as well as the internal management of all issues related to Security & Management of Sensitive & Personal Data within the company and do not concern the systems, data and processes of its customers.

Pseudonymization and Encryption of Personal Data

CSCY processes any personal data located within its infrastructure in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as such additional information is kept separately and is subject to appropriate technical and organizational measures such as to ensure that it cannot be attributed to an identified or identifiable natural person. Furthermore, CSCY encrypts personal data with symmetric and asymmetric keys.

Ensuring the security of workstations & communication network

- Company employees cannot disable or bypass security settings.
- Company employees are not able to install unauthorized software applications. The system has a session time-out system when the user is not active for a certain period of time.
- Antivirus installation takes place. Protection against all kinds of malware has been provided:
    - viruses, worms, trojan horses, rootkits, spyware and adware. More specifically: PC memory, executable files, protected and hidden files, removable storage media (CDs / DVDs / USB devices), incoming and outgoing network traffic of the carrier are checked
    - Checks are carried out in real time and not at the request of the user Malware and related files are automatically removed.
- Firewall installation takes place with a clearly defined inbound/outbound traffic policy.
- Intrusion detection/prevention mechanism (IDS/IPS) is implemented on critical systems.
- Antivirus applications and detection signatures are automatically updated.
- Wireless access to the IT system is allowed only to specific, authorized users and procedures.

Ensuring on an ongoing basis the integrity and availability of the data processed by the Company and the availability and reliability of its systems and services

© 2023-2025 BIO-STREAMS

- Data backup and restore procedures are clearly defined, documented and clearly linked to the roles and responsibilities of the relevant employees of the company.
- Backups provide a consistent level of physical and environmental protection with the standards applicable to the data at the time of its creation.
- The execution of backups is monitored to ensure their completion.
- Full backups are created regularly.
- Backups are safely stored in different locations.
- The mail servers used by the company provide protection:
    - Blocking messages considered spam (for example, using a blacklist of unwanted websites or mailing list servers).
    - Checking the integrity of messages for completeness of their content (for example, checking that a message includes all necessary headers.)
    - Don't automatically forward emails to external recipients.

Ensuring security of access to personal data

- The company ensures restriction of access of those who do not have the right to the points where personal data is kept (eg locked offices, cabinets, computer room with access control).
- The company ensures clear identification, through appropriate means, e.g. ID Badges, for all staff and visitors with access to the premises.
- The company ensures secure storage of files containing personal data (physical file). The physical archive is placed in cabinets, placed in places not exposed to public view.

Control procedures to ensure security of processing

- Information Security Management System
    - CSCY complies with the requirements of ISO 27001:2013 (Information Security Management System). To that end, it shall maintain an audit procedure based on the risk management approach based on that standard.
- Manage resources and systems
    - IT resources are reviewed and updated on a regular basis.
    - IT asset management processes are reviewed and updated annually.
    - Software development takes place in a special environment, which is not connected to the IT system used to process data. When testing is required, virtual data (not real data) is used.

The company is fully in line with the requirements of the European Regulation 2016/679 (General Data Protection Regulation, GDPR) (hereinafter: the Regulation) concerning the protection of personal data, fulfilling all its obligations arising from it. In this context, it has done the following:

Designation of a Data Protection Officer (DPO)

The company has appointed a Data Protection Officer who has the following tasks:

- informs and advises the company and its employees who process such data, about their obligations under the Regulation and other provisions of Greek and EU law on the protection of personal data,
- coordinates the cooperation between the company's employees in order to comply with the company's Regulation,

© 2023-2025 BIO-STREAMS

- monitors and supervises the company's compliance with the Regulation and the provisions of Greek and EU law on the protection of personal data,
- proposes the measures it deems necessary to achieve continuous compliance of the company with the requirements of the Regulation and recommends to the Management the appropriate policies, methodologies and practices for the processing of such data in order to protect them,
- informs the company's management about the progress of its compliance with the requirements of the Regulation,
- provide advice, when requested, on the data protection impact assessment and monitor its implementation in accordance with Rule 35 of the Regulation;
- consults with the Personal Data Protection Authority, in accordance with Article 36 of the Regulation,
- cooperates with the National Supervisory Authority, i.e. the Personal Data Protection Authority and represents the company vis-à-vis the National and European Personal Data Protection Authorities, in accordance with the Regulation, as responsible for the company's compliance with personal data protection legislation,
- acts as a contact point for the Personal Data Protection Authority on issues related to processing and will consult in accordance with article 36 of the Regulation, as appropriate, on any other matter.
- manages data security breaches that may arise in terms of how to deal with them and coordinates the Company's designated incident response team.

Subcontractors

- The standard guidelines and procedures regarding the processing of personal data by subcontractors with whom the company may cooperate are defined, documented and agreed through a data processing contract and in accordance with the Regulation
- Formal requirements and obligations are formally agreed between the company and any subcontractor.
- The company regularly checks the compliance of the subcontractor with the agreed level of requirements and obligations
- Employees of the subcontractor who process such data are subject to specific documented confidentiality/non-disclosure agreements.

Duty of confidentiality of staff

- Before assuming their duties, employees are informed about the company's security policy and are required to sign confidentiality agreements and non-disclosure of personal data lawfully processed in the context of their duties.
- The company ensures that all employees understand their responsibilities and obligations regarding the use of Data. The company's staff is informed about their roles and responsibilities in a complete and clear manner.

Security policy and procedures for the protection of personal data

The Company prepares and adopts manuals with the Policies and procedures for the security of the information it processes concerning:

- (the responsibilities and obligations of the Company and the Company's personnel regarding the lawful processing of Personal Data,
- the responsibilities and obligations of third parties cooperating with the Company regarding the lawful processing of Personal Data,
- the technical and organizational measures adopted by the Company for the safe processing of Personal Data.

© 2023-2025 BIO-STREAMS

Staff training

- The company ensures that all employees are adequately informed about IT security controls related to their daily work. Employees who lawfully process personal data on the basis of their duties are duly informed about the requirements for the protection of personal data provided by law, at the company's initiative, through regular information campaigns. To this end, the company regularly organises relevant training programmes for staff.

The company is actively pursuing the regular review, assessment and evaluation of the effectiveness of the technical and organizational measures applied to ensure the security of processing. This ensures the protection of the information to be processed, the applications, the operating environment and the technical application of protection concepts.

9.4 Data Processor: Novelcore (NVCR)

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Development of a synthetic data generator for BIOSTREAMS project that produces custom datasets on-demand that mimic actual data with no relation to any actual case.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: ……

SERVICES DESCRIPTION

The Synthetic Data Generator component will produce custom datasets on-demand. These models are trained offline on real data.

© 2023-2025 BIO-STREAMS

INTENDED PROCESSING ACTIVITIES

- ☐ Collection
- ☐ Registration
- ☑ Organisation
- ☑ Structuring
- ☐ Conservation
- ☑ Adaptation or modification
- ☐ Extraction
- ☐ Consultation
- ☑ Use
- ☐ Distribution or any other form of making available
- ☐ Rapprochement or interconnection
- ☐ Limitation
- ☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

- ☑ Patients (specify the medical services/units from which the patients will be recruited)
- ☑ Clinical Data Subjects
- ☑ Minors
- ☐ Current personnel
- ☐ Former personnel
- ☐ Volunteers
- ☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

- ☑ Data concerning health:
- ☐ Data revealing trade union membership
- ☑ Data revealing racial or ethnic origin
- ☐ Data revealing political opinions
- ☐ Data revealing religious or philosophical beliefs
- ☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

© 2023-2025 BIO-STREAMS

Please describe which data elements are in scope:

- Behavioural data: sleep duration, drink, eating patterns, physical activity
- Medical data: respiratory rate, blood pressure, asthma, BMI, BSA, liver fat measurements, breastfeeding duration, onset of cardiovascular disease
- Family data: cholesterol, BMI, history of diabetes, history of cardiovascular disease, blood pressure , history of hypertension

PLACE OF THE PROCESSING

Novelcore's premises in Athens and Patras in Greece.

THIRD PARTIES

N/A

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Dimitrios Tsakalidis tsakalidis@novelcore.eu	Dimitrios Tsakalidis tsakalidis@novelcore.eu

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities
- Information security policy
- Access management: 2-factor authentication for example (something you have + password + login)
- Directory of those processing personal data
- Process for regularly testing, assessing and evaluating
- Privileged identity management: minimise access to personal data
- Logging
- Disaster recovery plan
- Procedure for managing security incidents and data breaches
- Paper document security

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

9.5 Data Processor: AINIGMA Technologies (AINIGMA)

Annex I – Instructions and Processing Details

© 2023-2025 BIO-STREAMS

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Development and validation of two tools for the BioStreams project:

a. Recommendation Engine: A rule-based model (recommendation engine) for the delivery of personalised programmes of validated lifestyle recommendations for prevention and healthy living.
b. Risk Assessment: A ML-based model for the personalised prediction of the risk level of adverse metabolic outcomes attributed to obesity and overweight.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: .....

SERVICES DESCRIPTION

AINIGMA will perform data analysis, interpretation, and modelling for the BIO-STREAMS project. This includes the development (training & deploying) two AI/ML algorithms for the purposes of the Open Toolkit of the Active Health App. AINIGMA will also contribute to the communication and dissemination of research findings.

INTENDED PROCESSING ACTIVITIES

☐ Collection
☐ Registration
☑ Organisation
☑ Structuring
☐ Conservation
☑ Adaptation or modification
☐ Extraction
☑ Consultation
☑ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

☑ Patients (specify the medical services/units from which the patients will be recruited)

© 2023-2025 BIO-STREAMS

☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

☑ Data concerning health:

☐ Data revealing trade union membership

☑ Data revealing racial or ethnic origin

☐ Data revealing political opinions

☐ Data revealing religious or philosophical beliefs

☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

- Demographic data: sex, age, education, place of residence
- Lifestyle data: physical activity, sleep duration, eating and drinking patterns

PLACE OF THE PROCESSING

- AINIGMA Technologies, Innovation & Incubation Center KU Leuven, Kapeldreef 60, Leuven 3000, Belgium
- AINIGMA Technologies, Greek Branch, Ektoros 20, Gerakas Attikis 15334, Greece

THIRD PARTIES

Approved Subprocessor	Subprocessor location	Locations where In-Scope Personal data will be stored	Locations from which In-Scope Personal data will be accessed	Type of processing

© 2023-2025 BIO-STREAMS

Vlaams Supercomputer Centrum (VSC)	Datacenters of the universities of Antwerp, Brussels, Ghent and Leuven, Belgium	Ghent, Belgium	Not Applicable, Only used as a storing facility	Use of data for training AI models on the supercomputer

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Athanasios Kakasis kakasisathan@ainigma.tech	Athanasios Kakasis kakasisathan@ainigma.tech

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Records of processing activities
- Directory of those processing personal data
- Privileged identity management: minimise access to personal data
- Logging

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

© 2023-2025 BIO-STREAMS

9.6 Data Processor: HAROKOPIO UNIVERSITY OF ATHENS (HUA)

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Development of a serious games mobile application for BIOSTREAMS to support the prevention and treatment of childhood obesity through knowledge-based and cognitive tasks.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: .....

SERVICES DESCRIPTION

The serious games application will collect personal data from users, including demographic information (age, gender) and in-game behavior data (time spent, levels completed). It will also implement gamification techniques such as rewards, challenges, and progress tracking, as well as personalized feedback and suggestions to motivate users to consistently engage with the application and adopt healthier habits.

INTENDED PROCESSING ACTIVITIES

☐ Collection
☐ Registration
☑ Organisation
☑ Structuring
☐ Conservation
☑ Adaptation or modification
☐ Extraction
☑ Consultation
☑ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

© 2023-2025 BIO-STREAMS

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

- ☑ Patients (specify the medical services/units from which the patients will be recruited)
- ☑ Clinical Data Subjects
- ☑ Minors
- ☐ Current personnel
- ☐ Former personnel
- ☐ Volunteers
- ☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

- ☑ Data concerning health:
- ☐ Data revealing trade union membership
- ☑ Data revealing racial or ethnic origin
- ☐ Data revealing political opinions
- ☐ Data revealing religious or philosophical beliefs
- ☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

- Demographic data (age, gender) and
- In-game behavior data (time spent, levels completed)

PLACE OF THE PROCESSING

Harokopio University, Department of Informatics & Telematics 9, Omirou Str. 177 78, Tavros.

THIRD PARTIES

N/A

© 2023-2025 BIO-STREAMS

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME: DPO:	
For the Data Processor	Elena Politi politie@hua.gr	Elena Politi politie@hua.gr

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities
- Information security policy
- Access management: 2-factor authentication for example (something you have + password + login)
- Directory of those processing personal data
- Process for regularly testing, assessing and evaluating
- Privileged identity management: minimise access to personal data
- Logging
- Disaster recovery plan
- Procedure for managing security incidents and data breaches
- Paper document security

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

© 2023-2025 BIO-STREAMS

9.7 Data Processor: Innovation to Grow s.r.l (I2G)

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Identification of cost related data to incorporate cost-related indicators in an health economic model on overweight and obesity.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: ……

SERVICES DESCRIPTION

Retrospective and prospective data will be analysed to identify relations between variables to calibrate the health economic model implemented. These Cost data and selected behavioural and clinical data will be used to complement or update the literature on health economic indicators to be included in the health economic model on obesity and overweight developed by i2g in the project.

INTENDED PROCESSING ACTIVITIES

☐ Collection
☐ Registration
☑ Organisation
☑ Structuring
☐ Conservation
☐ Adaptation or modification
☐ Extraction
☑ Consultation
☑ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

© 2023-2025 BIO-STREAMS

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

☑ Data concerning health:

☐ Data revealing trade union membership

☑ Data revealing racial or ethnic origin

☐ Data revealing political opinions

☐ Data revealing religious or philosophical beliefs

☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

Economic relevant data:

- Cost of obesity-related hospital admissions directly attributable to obesity
- Cost of obesity-related hospital admissions with obesity as a secondary factor
- Cost of obesity-related outpatients activities (excluding rehabilitation)
- Cost of obesity-related drugs
- Cost of obesity-related rehabilitation services
- Total direct medical costs
- General practitioners obesity related visits
- Emergency department obesity related visits
- Caregivers' absenteeism due to obesity related assistance

Behavioural and clinical data:

- Body-mass index (BMI) (mean value)
- BMI z-score (mean value)
- Body-surface area (BSA) (mean value)

© 2023-2025 BIO-STREAMS

- Abdominal circumference (mean value)
- Percent body fat (mean value)
- Intake of dietary energy (daily mean kilojoules)
- Saturated fat intake (daily mean value)
- Physical activity (daily mean minutes)
  Moderate to vigorous physical activity (daily mean minutes)
  Time spent on outdoor activities (daily mean minutes)

PLACE OF THE PROCESSING

i2Grow srl, Via G. Matteotti, 21 40129 Bologna Italy

THIRD PARTIES

N/A

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Matteo Colombo m.colombo@i2grow.it	Matteo Colombo m.colombo@i2grow.it

© 2023-2025 BIO-STREAMS

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Backups
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities
- Information security policy
- Access management: 2-factor authentication for example (something you have + password + login)
- Directory of those processing personal data
- Process for regularly testing, assessing and evaluating
- Logging
- Disaster recovery plan
- Procedure for managing security incidents and data breaches

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

9.8 Data Processor: UKEMED GLOBAL Ltd (UKEMED)

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

© 2023-2025 BIO-STREAMS

☑ Scientific research: Study Oversight and Management, Risk Management, Data Analysis and Interpretation, Reporting and Publications

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: .....

SERVICES DESCRIPTION

UKEMED will perform data analysis, interpretation, and reporting for the BIO-STREAMS project. This includes overseeing data collection, managing risk, and ensuring data quality throughout the project lifecycle. UKEMED will also contribute to the publication of research findings.

INTENDED PROCESSING ACTIVITIES

☑ Collection
☑ Registration
☑ Organisation
☐ Structuring
☑ Conservation
☑ Adaptation or modification
☑ Extraction
☑ Consultation
☑ Use
☑ Distribution or any other form of making available
☑ Rapprochement or interconnection
☑ Limitation
☑ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☑ Volunteers
☐ Other (specify):

© 2023-2025 BIO-STREAMS

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

- ☑ Data concerning health:
- ☐ Data revealing trade union membership
- ☑ Data revealing racial or ethnic origin
- ☐ Data revealing political opinions
- ☐ Data revealing religious or philosophical beliefs
- ☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

- Biological data
- Demographic data
- Epigenetic data
- Health related data

PLACE OF THE PROCESSING

Data will be processed within the secure infrastructure of UKEMED, located in Cyprus. Managed computers connected to the local secured network and cloud storage solutions will be utilized to ensure data security and integrity.

THIRD PARTIES

N/A

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Takis Kotis	Takis Kotis

© 2023-2025 BIO-STREAMS

	takis@ukemed.com 121, Prodromou Street Offices 713-715 2064, Nicosia, Cyprus	takis@ukemed.com 121, Prodromou Street Offices 713-715 2064, Nicosia, Cyprus

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities
- Information security policy
- Access management: 2-factor authentication for example (something you have + password + login)
- Directory of those processing personal data
- Process for regularly testing, assessing and evaluating
- Privileged identity management: minimise access to personal data
- Logging
- Disaster recovery plan
- Procedure for managing security incidents and data breaches
- Paper document security

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

© 2023-2025 BIO-STREAMS

9.9 Data Processor: TECREANDO B.V. (TCR)

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☑ Scientific research: Development and training of ML tools for prognostic modeling, analytics, and personalised interventions.

☐ Other: (specify)

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: .....

SERVICES DESCRIPTION

TCR will support and contribute to the data analysis and prognostic modelling for the BIO-STREAMS project. This includes the development and training of AI/ML algorithms for the purposes of the Open Toolkit of the Active Health App. TCR will also contribute to the communication and dissemination of research findings.

INTENDED PROCESSING ACTIVITIES

☐ Collection
☐ Registration
☑ Organisation
☑ Structuring
☐ Conservation
☑ Adaptation or modification
☐ Extraction
☑ Consultation
☑ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

© 2023-2025 BIO-STREAMS

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☐ Other (specify):

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

☑ Data concerning health:

☐ Data revealing trade union membership

☑ Data revealing racial or ethnic origin

☐ Data revealing political opinions

☐ Data revealing religious or philosophical beliefs

☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

Please describe which data elements are in scope:

- Demographic data: sex, age, education, place of residence
- Lifestyle data: physical activity, sleep duration, eating and drinking patterns

PLACE OF THE PROCESSING

- TECREANDO B.V. Amsterdam Office, Herengracht 575-577, 1017 CD, Amsterdam, Netherlands
- TECREANDO B.V. Athens Office, Char. Trikoupi 18, 5th Floor, 10679, Athens, Greece

THIRD PARTIES

N/A

© 2023-2025 BIO-STREAMS

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Mr. Panagiotis Tsatsoulis TECREANDO B.V. Athens Office, Char. Trikoupi 18, 5th Floor, GR-10679, Athens, Greece tsatsoulis@tecreando.com Tel.: +30 2114441694 Mob.: +30 6942462881	Mr. Panagiotis Tsatsoulis TECREANDO B.V. Athens Office, Char. Trikoupi 18, 5th Floor, GR-10679, Athens, Greece tsatsoulis@tecreando.com Tel.: +30 2114441694 Mob. +30 6942462881

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities

© 2023-2025 BIO-STREAMS

- Directory of those processing personal data
- Privileged identity management: minimise access to personal data
- Logging

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

9.10 Data Processor: EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLGISTON-EMP (ICCS)

Annex I – Instructions and Processing Details

NATURE AND PURPOSE OF THE INTENDED PROCESSING OF PERSONAL DATA

Use the checkboxes to specify the purpose(s) of the processing that is/are considered in this agreement:

☐ Scientific research:

☑ Other: Hosting services

If treatment activities are conducted for research purposes, please check the appropriate box:

☑ Retrospective research
☑ Prospective research

Protocol number: .....

SERVICES DESCRIPTION

ICCS only provides hosting services.

INTENDED PROCESSING ACTIVITIES

☐ Collection
☑ Registration
☑ Organisation

© 2023-2025 BIO-STREAMS

☐ Structuring
☑ Conservation
☐ Adaptation or modification
☐ Extraction
☐ Consultation
☐ Use
☐ Distribution or any other form of making available
☐ Rapprochement or interconnection
☐ Limitation
☐ Erasing or destroying

CATEGORIES OF DATA SUBJECTS

Use the checkboxes to specify the categories of data subjects that are considered in this agreement:

☑ Patients (specify the medical services/units from which the patients will be recruited)
☑ Clinical Data Subjects
☑ Parents (Informal carers)
☑ Minors
☐ Current personnel
☐ Former personnel
☐ Volunteers
☑ Other (specify): Policy Makers, Tutors

TYPE OF PERSONAL DATA

Use the checkboxes to specify the categories of personal data that are considered in this agreement:

☑ Sensitive personal data:

☑ Data concerning health:

☐ Data revealing trade union membership

☑ Data revealing racial or ethnic origin

☐ Data revealing political opinions

☐ Data revealing religious or philosophical beliefs

☐ Data concerning sex life or sexual orientation

☐ Other personal data (not mentioned above):

© 2023-2025 BIO-STREAMS

PLACE OF THE PROCESSING

The services developed by the partners of BIOSTREAMS project are hosted on servers owned by the Biomedical Engineering Laboratory. The servers are physically located at the server room of the Laboratory in the University Campus of the National Technical University of Athens and are only accessible by authorized personnel in a locked room. The NTUA Zografou campus address is: 9, Heroon Polytechneiou Avenue, 15772.

THIRD PARTIES

Approved Subprocessor	Subprocessor location	Locations where In-Scope Personal data will be stored	Locations from which In-Scope Personal data will be accessed	Type of processing
TELEMATIC MEDICAL APPLICATIONS EMPORIA KAI ANAPTIXI PROIONTON TILIATRIKIS ETAIRIA PERIORISMENIS EYTHINIS (TMA)	ALEXANDROU PAPANASTASIOU AVE 151, PIRAEUS 185 33, Greece	Data related to nutritional and lifestyle markers will be collected by the ActiveHealth App. No other personal data will be stored neither in the ActiveHealth App, nor in the Dashboard.	Data related to the use of the Open Toolkit and other components will be forwarded by the Activehealth App to the relevant tools. No personal data will be viewed in the Dashboard.	Collection of data related to nutritional and lifestyle markers and processing required to view the data in the Dashboard.
NETCOMPANY-INTRASOFT SA	RUE NICOLAS BOVE 2B, LUXEMBOURG 1253, Luxembourg	Biobank IMS (Information Management System) will incorporate data from retrospective and prospective cohorts	No personal data will be viewed in the IMS.	No personal data processing

© 2023-2025 BIO-STREAMS

CONTACT DETAILS IN CASE OF PRIVACY REQUESTS (E.G.DPO) AND DATA BREACHES

	Privacy contact	**Data breaches contact**
For the Data Controller	NAME DPO:	
For the Data Processor	Dimitrios Kalogeras (DPO)	Dimitrios Kalogeras (DPO)

Annex 2 – Technical and Organisation Security Measures

The Data Processors should provide sufficient guarantees, in particular in terms of expert knowledge, reliability and resources, to implement technical and organisational measures as mentioned in article 32 GDPR which will meet the requirements of the GDPR, including for the security of processing.

In order to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Data Processors shall implement appropriate technical and organisational measures.

Technical measures could be one of the following:

- Pseudonymization and encryption of personal data and encryption of communications in case of transfer
- Firewall
- Anti-malware software
- Backups
- Extra servers in case needed
- Network infrastructure and workstation security
- Connection traceability (log file)
- User authentication

Organisational measures could be one of the following

- Physical access control
- Records of processing activities
- Information security policy
- Access management: 2-factor authentication for example (something you have + password + login)
- Directory of those processing personal data
- Process for regularly testing, assessing and evaluating
- Privileged identity management: minimise access to personal data
- Logging
- Disaster recovery plan
- Procedure for managing security incidents and data breaches
- Paper document security

© 2023-2025 BIO-STREAMS

When assessing the appropriate level of security, the risks that are presented by processing should be taken into account, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

© 2023-2025 BIO-STREAMS

10 Appendix D: Standard Operating Procedures for the Retrospective Study

Standard Operating Procedures for the Retrospective Study:

Data Monitoring and Adverse Effects Management

Revision: v.3.0

Work package	WP6
Task	Task 6.3
Document lead	UKCM, UM
Version	3.0
Authors	Izidor Mlakar

Co-funded by the European Union

© 2023-2025 BIO-STREAMS

Abstract	This document establishes a comprehensive framework for monitoring data-related adverse effects within the BIO-STREAMS consortium. The protocol is designed as a necessary activity to support execution of Retrospective Study (registered at https://doi.org/10.1186/ISRCTN12357025). It implements a multi-tiered monitoring structure that integrates expertise from clinical, technical, and ethical domains through a carefully designed committee system. The core committee includes The Coordinator (ICCS), Clinical Manager (UM), and Regulatory & Ethics (WLC), and Scientific and Technical Manager (HUA). The committee is further supported by PIs of the clinical sites and External Advisory Board of international experts in nutrition, pediatrics, and global health. This document details Standard Operating Procedures (SOPs) for data access monitoring, transfer supervision, and security incident management, while establishing clear reporting hierarchies and decision-making processes.
Keywords	standard operating procedure, adverse effect monitoring, data management

DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
1.0	04.12.2024	Initial draft of the deliverable	Izidor Mlakar
2.0	18.12.2024	Refined Version	All Partners
3.0	19.02.2025	Final version	All Partners

Disclaimer

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the other granting authorities. Neither the European Union nor the granting authority can be held responsible for them.

Copyright notice

© 2023 - 2025 BIO-STREAMS Consortium

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the document	Standard Operating Procedure	
Dissemination Level		
	SENSISTIVE, CONFIDENTIAL	X

10.1 PART A: GENERAL INFORMATION

10.1.1 Purpose of the Protocol

The BIO-STREAMS consortium represents a complex international collaboration in health research, bringing together diverse expertise across multiple domains. In this context, the need for robust data monitoring becomes paramount, not merely as a regulatory requirement but as a fundamental enabler of scientific progress and ethical research conduct.

This protocol aims to establish a framework for monitoring adverse effects associated with data breaches and leaks in the context of the BIO-STREAMS Retrospective Study, registered at https://doi.org/10.1186/ISRCTN12357025. It provides guidelines for identifying, reporting, and mitigating risks related to unauthorized access and misuse of sensitive data, ensuring the protection of individuals' privacy and the integrity of research data.

With this protocol for monitoring of data operations, we create an environment where researchers can confidently share and utilize data while maintaining the highest standards of data protection and ethical conduct.

10.1.2 Scope of the Protocol

The protocol's scope is not covering only the technical aspects of data monitoring, but also the human and organizational dimensions that enable implementation of an effective oversight. This protocol applies to all personnel involved in the handling, processing, and transfer of sensitive data related to retrospective studies. It encompasses all forms of data breaches and leaks, including but not limited to:

- Unauthorized access by internal or external parties.
- Accidental exposure of confidential information.
- Data loss due to technical failures or human error

10.1.3 Definitions

Adverse Effect: Any negative impact resulting from a data breach or leak, including harm to individuals, reputational damage to organizations, legal consequences, and financial loss.

Data Breach: An incident where unauthorized access to sensitive data occurs, potentially compromising confidentiality and security.

Data Leak: An unintentional release of confidential information that may expose sensitive data to unauthorized parties.

Sensitive Data: Any information that, if disclosed without authorization, could cause harm or distress to individuals or organizations. This includes personally identifiable information (PII), health records, financial information, and proprietary research data.

© 2023-2025 BIO-STREAMS

10.1.4 Roles and Responsibilities

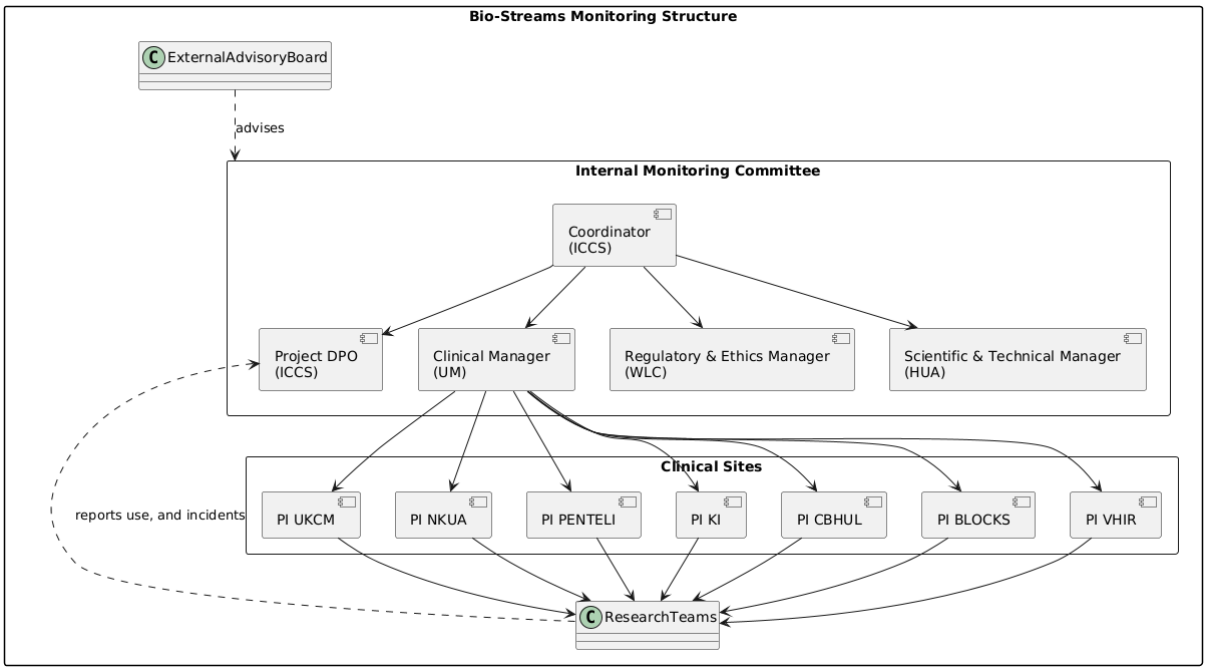

Figure 4: Structure in BIO-STREAMS

Project's Data Protection Officer (DPO): Responsible for overseeing compliance with data protection regulations, conducting risk assessments, and coordinating incident response efforts.

Research Team Members: Required to adhere to the protocol's guidelines for data handling and reporting any suspected breaches or leaks immediately. This includes technical personnel implementing technical safeguards, monitoring systems for unauthorized access, and maintaining logs of data transfers (on the targeted systems).

The BIO-STREAMS Ethics Advisory Board (EAB): Consists of the Coordinator (ICCS), Project's Data Protection Officer (ICCS), Clinical Manager (UM), Regulatory & Ethics Manager (WLC), and Scientific and Technical Manager (HUA). The committee is further supported by Principal Investigators of the clinical sites (UKCM, NKUA, PENTELI, KI, CBHUL, BLOCKS and VHIR)

The **Coordinator's role** as committee chair ensures that monitoring processes remains practical and implementable, while the **Scientific and Technical Manager's** and **Clinical Manager's** oversight guarantees that these processes serve rather than hinder research objectives. The **Regulatory & Ethics Lead's** involvement ensures that all monitoring activities align with ethical principles and legal requirements.

The **Principal Investigators (PIs)** at clinical sites function as main figures in the data monitoring framework, ensuring local data integrity and establishing links in the consortium's broader data oversight chain. Their role encompasses multiple layers of responsibility, starting with the fundamental oversight of data quality and completeness at their respective sites. The PIs serve as the first line of defense in data protection, taking immediate action when incidents occur and implementing necessary corrective measures, while maintaining clear communication channels with both their local teams and the BIO-STREAMS SOP committee. Beyond their local oversight, PIs participate actively in the consortium's broader monitoring structure through regular committee meetings, status reporting, and direct collaboration.

External Advisory Board members actively contribute to shaping monitoring practices that reflect real-world research needs and challenges. Their involvement ensures that monitoring

© 2023-2025 BIO-STREAMS

procedures remain relevant to the scientific goals while meeting international standards of research governance.

Table 5: Persons involved in the BIO-STREAMS Regulatory & Ethics Committee

Role	Name and surname	Institution	e-mail
CO	Eleftheria Vellidou	ICCS	ebel@biomed.ntua.gr
DPO	Dimitrios Kalogeras	ICCS	dkalog@gmail.com
CM	Izidor Mlakar	UM	izidor.mlakar@um.si
R&E	Peter Davis	WLC	pda@whitelabelconsultancy.com
S&T	Eleni Politi	HUA	politie@hua.gr
PI-BLOCKS	Radka Savova	BLOCKS	savova.radka@gmail.com
PI-CHUL	Caroline Gernay	CHUL	Cgernay@chuliege.be
PI-KI	Billy Langlet	KI	billy.langlet@ki.se
PI-NKUA	Evangelia Charmandari	NKUA	evangelia.charmandari@googlemail.com
PI-PENTELI	Olga Fafoula	PENTELI	ofafoula@gmail.com
PI-UKCM	Martin Bigec	UKCM	Martin.bigec@ukc-mb.si
PI-VHIR	Andreea Ciudin	VHIR	andreea.ciudin@vhir.org

* Note: CO – Coordinator, DPO - Project's Data Protection Officer, CM - Clinical Manager, R&E - Regulatory & Ethics Manager, S&T - Scientific and Technical Manager, PI - Principal Investigator

10.1.5 Regulatory Compliance

At the foundation of regulatory compliance framework within the BIO-STREAMS the General Data Protection Regulation (GDPR) pilar, which permeates all aspects of data handling within the consortium. The implementation of GDPR requirements extends beyond mere technical compliance, encompassing a comprehensive understanding of data protection principles and their practical application in research settings. Each participating institution maintains dedicated Data Protection Officers who coordinate their efforts to ensure consistent application of GDPR principles across the consortium while adapting to local regulatory landscapes.

The oversight of research ethics and protocol compliance is the second crucial pillar of our regulatory framework. Each clinical site operates under approved protocols from their respective Institutional Review Boards (IRBs) or Ethics Committees, with these approvals forming the backbone of our research conduct and access to data. The consortium maintains

© 2023-2025 BIO-STREAMS

a centralized repository of all IRB approvals, protocol amendments, and related documentation, ensuring transparency and facilitating regular compliance audits. This systematic approach to protocol management enables quick identification and resolution of potential compliance issues while maintaining the highest standards of research ethics.

Data processing and transfer within the consortium operate under a framework of agreements that define responsibilities and ensure compliance with both GDPR and IRB requirements. The Data Processing Agreement (DPA) serves as the baseline document, establishing common ground rules for all consortium members.

Incident monitoring and management is the final pilar of our regulatory compliance framework. The consortium maintains clear procedures for identifying, reporting, and responding to potential compliance breaches. These procedures ensure prompt notification of relevant authorities when required while facilitating quick implementation of corrective measures. The incident response framework, presented in the next sections of this document, includes documentation requirements and post-incident review procedures, enabling the consortium to learn from experiences and strengthen preventive measures.

10.2 PART B: OPERATIONAL PROCEDURES

The operational procedures outlined in this section serve as guidance for implementing the data monitoring framework across the BIO-STREAMS consortium. These procedures are set in place to ensure consistent, effective monitoring while maintaining flexibility to accommodate site-specific needs and varying data types.

10.2.1 Committee Operations

Committee operations serve as the primary mechanism for coordinating monitoring activities across the consortium, ensuring that all stakeholders remain aligned in their approach to data protection and oversight.

Regular Meetings

The meetings operate at different levels of frequency and formality, each serving specific purposes within the broader monitoring framework. The monthly general meetings focus on operational aspects and immediate concerns, while quarterly reviews provide opportunities for strategic assessment and long-term planning. Emergency meetings, while rare, ensure the consortium's ability to respond rapidly to urgent situations requiring immediate attention.

The following outlines the nature and structure of the meetings:

Monthly General Meetings

- Chair: Project Coordinator (ICCS)
- Required attendance: All BIO-STREAMS Regulatory & Ethics Committee members, Clinical Site PIs
- Duration: up to 2 hours
- Format: Virtual

Bi-Annual Review Meetings (during live consortium meetings)

- Chair: Project Coordinator (ICCS)
- Required attendance: BIO-STREAMS Regulatory & Ethics Committee, External Advisory Board, Clinical Site PIs, Researchers

© 2023-2025 BIO-STREAMS

- Duration: 2-3 hours
- Format: Hybrid with physical presence recommended

Emergency Meetings

- Triggered by: Security incidents, major protocol deviations, urgent regulatory matters
- Called by: Project DPO or Coordinator
- Required attendance: All BIO-STREAMS Regulatory & Ethics Committee members, Clinical Site PIs concerned, Researchers Concerned
- Maximum response time: 24 hours
- Format: Virtual for immediate response

The general meeting structure should follow the following example (can be adapted to the specific relevant issue):

1. Review of previous actions (15 minutes)
2. Site reports and updates (30 minutes)
3. Incident review if applicable (30 minutes)
4. Risk assessment and mitigation (30 minutes)
5. Action items and assignments (15 minutes)

10.2.2 Decision-Making Protocols

The BIO-STREAMS consortium implements a tiered decision-making framework that recognizes different categories of decisions and their respective impact on data monitoring activities. This structured approach ensures appropriate consideration and oversight while maintaining operational efficiency and the ability to respond to urgent situations.

Standard Decisions

Standard decisions encompass routine operational matters that affect day-to-day monitoring activities. These decisions require a simple majority vote of the BIO-STREAMS Regulatory & Ethics Committee, ensuring broad consensus while maintaining operational efficiency. For standard decisions, the implementation timeline must be clearly specified. Responsible parties must be explicitly identified, with clear delineation of roles and responsibilities in implementing the decision.

Each standard decision must be formally documented in the meeting minutes, including the specific points of discussion, voting outcomes, and any significant concerns raised during deliberation. Examples of standard decisions include:

- Modifications to routine monitoring schedules
- Updates to documentation templates
- Minor adjustments to reporting procedures
- Regular review and update of access permissions

Critical Decisions

Critical decisions involve significant changes to monitoring procedures or responses to substantial security concerns. These decisions require a higher threshold of approval, specifically a 75% majority of the BIO-STREAMS Regulatory & Ethics Committee (present at the meeting), including the PIs of the clinical sites, reflecting their greater impact and potential risks. A formal implementation plan for such decisions needs to include:

© 2023-2025 BIO-STREAMS

- Detailed timeline with specific milestones
- Resource allocation requirements
- Impact analysis on existing procedures
- Risk mitigation strategies
- Success metrics and evaluation criteria

The External Advisory Board should be consulted on these decisions, providing independent expert perspective and additional oversight. This consultation must be documented, including the Board's recommendations and any reservations expressed. A comprehensive written risk assessment must accompany each critical decision, analysing potential impacts on data security, operational efficiency, and regulatory compliance.

Emergency Decisions

Emergency decisions address urgent situations requiring immediate action to protect data security or maintain essential operations. These decisions can be made by a minimum of four committee members, necessarily including the Project Coordinator (or delegate) and Project DPO, plus at least two additional committee members. This streamlined decision-making process enables rapid response while maintaining essential oversight. Such decisions must be documented, including:

- Circumstances necessitating emergency action
- Details of the decision-making process
- Immediate actions taken
- Preliminary impact assessment
- Initial risk mitigation measures

All committee members must receive notification within 24 hours of an emergency decision, including:

- Complete description of the situation
- Actions taken
- Rationale for emergency handling
- Initial outcomes
- Next steps

The emergency decision must undergo review at the next regular meeting of the BIO-STREAMS Regulatory & Ethics Committee.

10.2.3 Reporting Framework

The reporting framework is designed to ensure effective communication, timely oversight, and continuous improvement of data monitoring activities.

Routine Reporting

Routine reporting forms the basis of the consortium's monitoring oversight, providing regular updates on operational status, compliance measures, and emerging concerns. It consists of:

1. **Monthly Reports -** provide detailed operational insights into monitoring activities across all consortium sites. These reports serve as the primary mechanism for tracking

© 2023-2025 BIO-STREAMS

routine monitoring activities and identifying emerging trends or concerns. The reports should include:

- Operational metrics from both clinical and technical sites,
- Data access and transfer statistics,
- Security status updates,
- Compliance verification results,
- Incident summaries and resolutions,
- Emerging risks or concerns, and
- Action items and their status

The routine reports are generated jointly by the clinical and technical teams and the Regulatory & Ethics Committee.

2. **Bi-Annual Reviews -** provide a broader perspective on monitoring effectiveness and emerging trends. These assessments synthesize information from monthly reports while incorporating additional analysis and strategic considerations to be discussed on the consortium level as a whole. Such reports should include:
    - Trend analysis across 6 months of monitoring data
    - Progress on long-term initiatives
    - Resource allocation effectiveness
    - Technology infrastructure assessment
    - Recommendations for system improvements

Incident Reporting

Incident reporting ensures rapid response to security events while maintaining clear documentation for reporting to external authorities, analysis and improvement. The process includes:

1. **Immediate Notifications** - required for any security incident or significant monitoring concern requiring prompt attention. Notifications must include:
    - Incident description and classification
    - Initial impact assessment
    - Immediate actions taken, including which authorities were notified and when
    - Resources required
    - Next steps planned
    - Contact information for responsible parties

2. **Follow-up Reports** - follow-up reports provide detailed analysis of incidents and response effectiveness. These must be submitted within 48 hours of incident resolution. Reports should include:
    - Incident timeline
    - Root cause analysis
    - Impact assessment
    - Resource utilization details
    - Lessons learned
    - Recommendations for prevention
3. **Resolution Documentation** - it provides complete records of incident handling and closure. This documentation serves both operational and compliance purposes. Thus, Key components include:
    - Final incident analysis
    - Complete audit log
    - All actions taken
    - Resources utilized

© 2023-2025 BIO-STREAMS

- Preventive measures implemented
- Compliance implications
- Sign-off from relevant authorities

Ethics Advisory Board Reporting

Ethics Advisory Board (EAB) reporting ensures independent oversight and strategic guidance for monitoring activities. This primarily includes **Consultation Procedures** on topics such as:

- Regular briefing schedules
- Emergency consultation protocols
- Documentation requirements
- Response timelines
- Implementation of the protocol and tracking
- Incident response procedures and impact assessment

10.3 PART C: Monitoring Procedures

The BIO-STREAMS monitoring procedures are adapted to the specific roles and responsibilities of clinical and technical partners. These procedures ensure secure and efficient data handling while maintaining clear accountability and oversight.

© 2023-2025 BIO-STREAMS

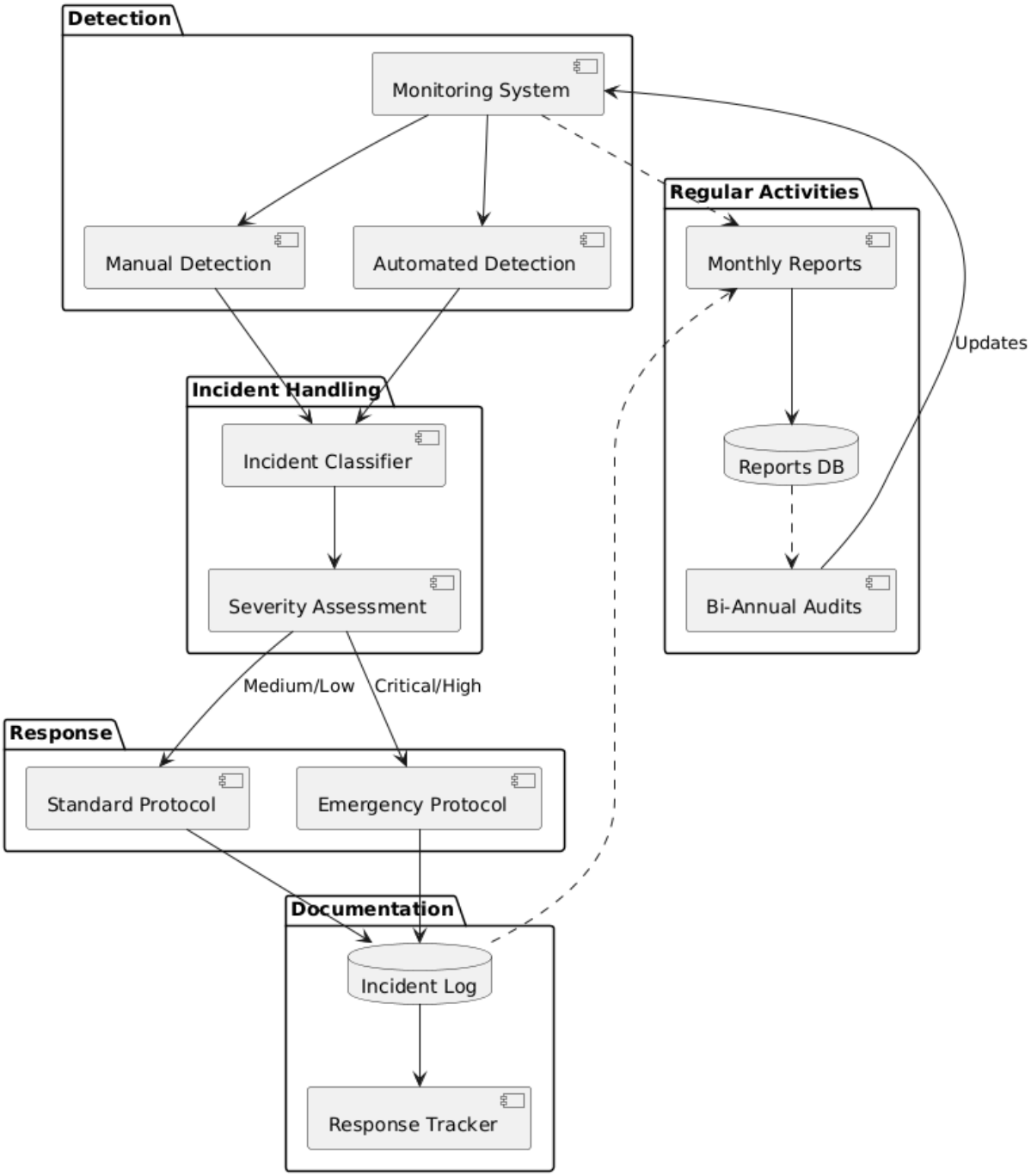

Figure 5: Monitoring Framework

10.3.1 Data Transfer Monitoring

Clinical partners are responsible for monitoring all aspects of data transfer from the point of collection to handover to the technical teams. This monitoring ensures data integrity, security, and compliance with protocol requirements.

Primary Responsibilities Include:

- Verification of data completeness and quality before transfer
- Documentation of transfer timing and volumes

© 2023-2025 BIO-STREAMS

- Validation of transfer security measures
- Tracking of data transformations and anonymization
- Monitoring of transfer authorization compliance

The monitoring activities include:

- Pre-Transfer Validation
- Transfer Process Monitoring

Pre-Transfer Validation

The clinical team must, in collaboration with WP3, perform systematic quality validation before any data transfer. This process includes verification of data completeness, accuracy, and consistency with established protocols. Teams must check that all required fields are populated, values fall within expected ranges, and any derived variables are correctly calculated. Quality assessment must be documented using standardized checklists, with any deviations noted and resolved before transfer authorization.

Clinical teams must validate that data adheres to the consortium's standardized format specifications. This includes verification of:

- File formats and structures
- Variable naming conventions
- Coding schemes
- Date formats
- Missing data codes
- Special character handling
- Version control elements

See Data Transfer & Quality Check Form (in Appendix).

Transfer Process Monitoring

Clinical teams must actively monitor the security of data transfer channels by verifying encryption protocols, checking if the connection is secure, validating authentication mechanisms and documenting security parameters.

Each transfer requires detailed documentation including:

- Transfer initiation time
- Completion time
- Sender and recipient details
- Security measures employed
- Verification procedures used
- Any issues encountered

See **Error! Reference source not found.** (in Appendix).

10.3.2 Data Access Monitoring

Data access monitoring is the essential responsibility of technical partners with access to the BIO-STREAMS retrospective data. The monitoring framework ensures the security, appropriate use, and traceability of all data access activities while maintaining research efficiency and regulatory compliance.

© 2023-2025 BIO-STREAMS

Access Control Monitoring

Technical teams must implement and maintain a robust authentication monitoring system that tracks all aspects of user access. This system operates continuously, providing real-time monitoring and alerting capabilities.

Authentication Tracking Requirements (related to infrastructure and systems that come in contact with retrospective data):

- Continuous monitoring of all login attempts to infrastructure that hosts the retrospective data, both successful and failed
- Recording of session metadata including duration, IP addresses, and device information
- Tracking of all authentication method changes and updates
- Monitoring of password policy compliance and change patterns
- Documentation of multi-factor authentication events
- Recording of session termination events, both user-initiated and automatic
- Tracking of concurrent session attempts

See **Error! Reference source not found.** (in Appendix).

Permission Management Framework

Technical partners must maintain a comprehensive permission management system that ensures appropriate access levels and maintains detailed audit trails for all systems that have access to retrospective data. Prior to getting access to data a report should include:

- Definition and maintenance of role hierarchies
- Description of approval workflows
- Description of authorization chains
- Description of how the assignments and changes to roles are tracked
- Description of how temporary role assignments are tracked

Each new role or assignment of a new user should be tracked and documented alongside any changes in access privileges. These should be turned to the monitoring board on monthly basis.

See **Error! Reference source not found.** (in Appendix).

Activity Monitoring System

Technical teams must implement activity monitoring that tracks all interactions with research data. The monitoring requirements include:

1. Data Access Patterns: frequency and timing, volume of data accessed, access duration patterns
2. Processing Activities: processing duration tracking, algorithm performance monitoring, data transformation logging, Query patterns, Output generation patterns

See **Error! Reference source not found.**(in Appendix).

© 2023-2025 BIO-STREAMS

10.3.3 Security Incident Monitoring

Security incident monitoring within the BIO-STREAMS consortium must encompass a multi-layered approach integrating both preventive and reactive measures. This critical function requires coordinated effort between clinical and technical partners to ensure comprehensive coverage of all potential security threats to research data. The monitoring framework operates continuously across all consortium systems, employing automated detection tools, manual reporting mechanisms, and regular security assessments. Incident monitoring described in this section needs to follow a structured approach beginning with detection through automated systems that provide real-time alerts for potential security breaches, unauthorized access attempts, or suspicious patterns of data usage.

Upon detection, incidents undergo immediate triage and classification based on severity, scope, and potential impact on data security and research operations. This classification determines the response protocol and escalation path, with critical incidents triggering immediate notification to the Relevant Pilot Site PI, Project DPO and relevant consortium leadership. The monitoring system maintains comprehensive audit trails of all security events, including attempted breaches, successful intrusions, and near-misses, enabling both immediate response and long-term pattern analysis.

Security Framework

Technical teams must implement security monitoring system for:

- Real-time security monitoring
- Access violation detection
- Data integrity checking
- Network security tracking
- Encryption status verification

Technical teams must conduct regular audits to ensure compliance with consortium policies and regulatory requirements. The Audit Components include:

- Access Control Audits (e.g. User account reviews, Permission assignment verification, Role configuration assessment, Authentication method validation)
- Usage Audits (e.g. Purpose limitation verification, Data minimization compliance, Processing restriction adherence)
- Security Control Audits (e.g. Encryption implementation, Access control effectiveness, Logging system completeness, Backup system verification)

© 2023-2025 BIO-STREAMS

Incident Assessment and Response

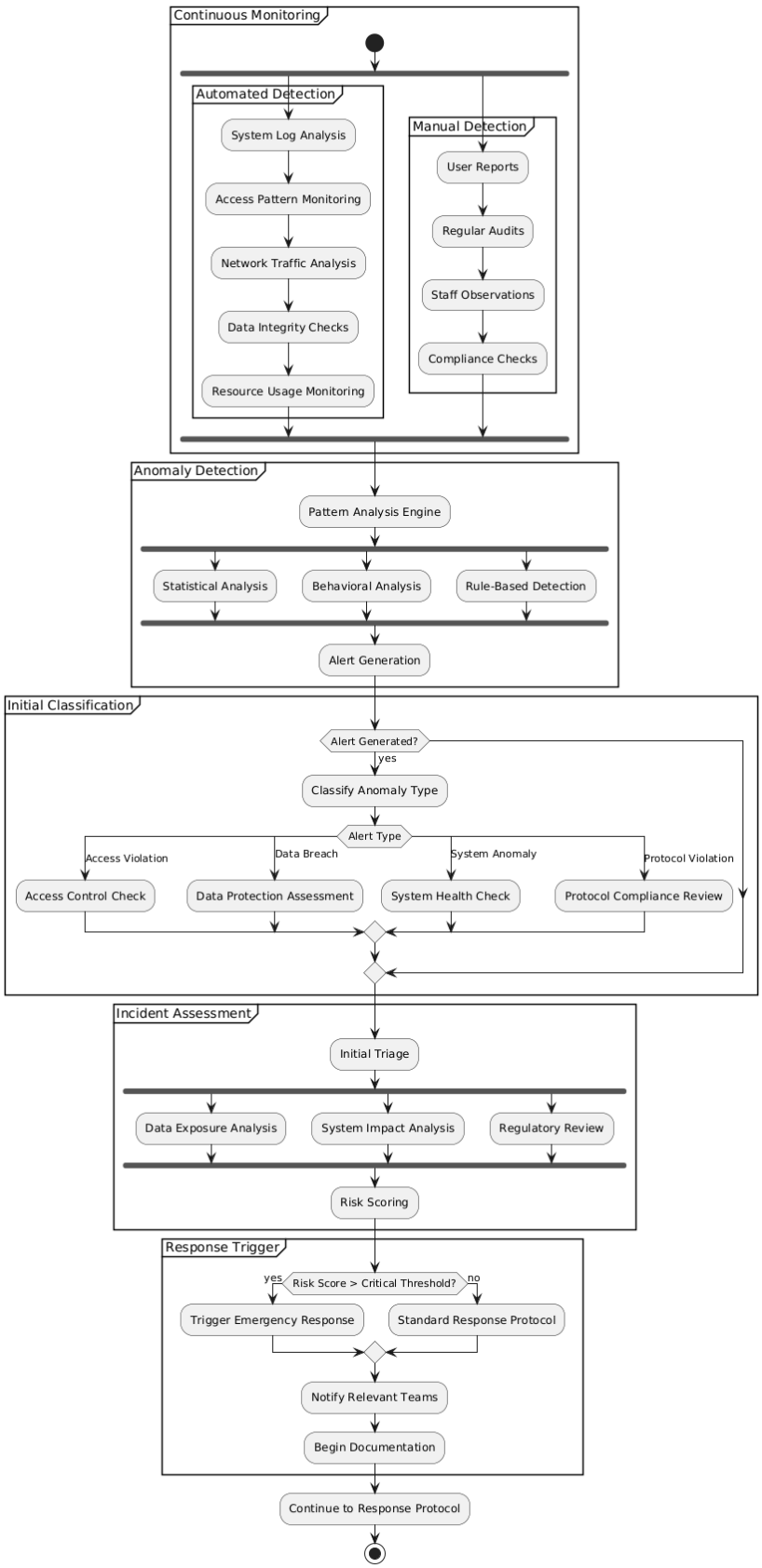

Figure 6: An example of BIO-STREAMS Security Framework's flow from Incident Detection to Response

© 2023-2025 BIO-STREAMS

If an incident is detected, it needs to undergo systematic assessment immediately upon detection. The assessment phase evaluates multiple dimensions of the incident: scope of potential data exposure, number of affected systems or datasets, geographical spread across consortium sites, and potential regulatory implications.

Technical teams must quantify immediate impact while clinical teams assess potential implications for patient data and research integrity. This initial assessment determines the incident severity classification and triggers appropriate response protocols.

Impact analysis examines both immediate and potential long-term consequences of the security incident. The analysis considers direct effects on data security, potential compromise of research integrity, and broader implications for consortium operations. Key considerations include:

- Extent of data exposure or corruption
- Number of affected participants or datasets
- Impact on ongoing research activities
- Potential regulatory compliance violations
- Resource requirements for remediation
- Reputational risks to the consortium

The Response coordination also follows a structured approach based on incident severity. For critical incidents immediate containment measures need to be implemented. The Project DPO assumes overall coordination, ensuring communication between clinical and technical teams. Response activities include:

- Implementation of containment measures
- Coordination of technical remediation efforts
- Management of communication channels
- Oversight of recovery operations
- Documentation of response actions

Every incident requires comprehensive documentation from detection through resolution. The documentation captures:

- Chronological incident timeline
- All response actions taken
- Resource allocation details
- Communication records
- Resolution milestones

10.3.4 Data Transfer & Quality Check Form

TRANSFER INFORMATION	
Transfer ID (SITE-YYYYMMDD-##)	
Date of Transfer	

© 2023-2025 BIO-STREAMS

(DD/MM/YYYY)	
Time of Transfer (24hr format)	
SOURCE INFORMATION	
Sending Site ID	_____
Sending Institution	_____
Responsible Person	_____
Contact Details	Email: _____ Phone: _____
DESTINATION INFORMATION	
Receiving Site ID	_____
Receiving Institution	_____
Responsible Person	_____
Contact Details	Email: _____ Phone: _____
DATA DESCRIPTION	
Data Type	□ Raw Clinical Data □ Harmonized Clinical Data □ Demographic Data □ Biochemical Data □ Lifestyle/Behavioral Data □ Imaging Data □ Other: _____
Cohort	□ Overweight/Obese (n= **)** □ *Normal Weight (n*= **)**
Age Range	_____ to _____ years
Time Period Covered	_____ to _____
File Format	□ CSV □ XML □ JSON □ SQL □ Other: _____
File Size	_____ MB/GB

© 2023-2025 BIO-STREAMS

Number of Records	_____
Data Identifiers	□ Fully Anonymized □ Pseudonymized □ Contains Identifiers: _____
TRANSFER METHOD	
Transfer Protocol	□ Secure FTP □ Encrypted Email □ Secure Cloud Storage □ Dedicated Research Network □ Physical Media: _____ □ Other: _____
Encryption Method	_____
Verification Hash	_____
COMPLIANCE	
Ethics Approval Reference	_____
Data Transfer Agreement	□ Yes (Ref: _____) □ No
GDPR Compliance Verified	□ Yes □ No (Reason: _____)
Data Protection Impact	□ Completed (Date:_____) □ Not Required (Reason: _____)
QUALITY CONTROL	
Data Completeness according to the BIO-STREAMS data model	□ Complete □ Incomplete (Details: _____)
Data Integrity	□ Verified □ Failed (Details: _____)
Missing Data Percentage	_____%

© 2023-2025 BIO-STREAMS

TRANSFER STATUS	
Transfer Outcome	□ Successful □ Partially Successful □ Failed (Reason: _____)
Receipt Confirmation	□ Confirmed (Date: _____) □ Pending □ Not Received
AUTHORIZATION	
Sender Signature & Date	_____ / _____
Receiver Signature & Date	_____ / _____
DPO Review	□ Approved □ Pending □ Rejected (Reason: _____)
DPO Signature & Date	_____ / _____

© 2023-2025 BIO-STREAMS