![Smart CHANGE logo]

**Empowering Youth with AI for healthier lives**

| Project Title | AI-based long-term health risk evaluation for driving behaviour change strategies in children and youth |
|---|---|
| Project Acronym | SmartCHANGE |
| Grant Agreement No. | 101080965 |
| Project Start Date: | 1 May 2023 |
| Project Duration: | 48 months |
| Project Website: | https://www.smart-change.eu/ |

# D1.2 – Data Management Plan

| Work Package | WP1 |
|---|---|
| Lead Partner | UPRC |
| Contributing Author(s)(Partner) | Anton Gradišek (JSI), José Carlos Ribeiro (UPORT), Thanos Kiourtis (UPRC), Argyro Mavrogiorgou (UPRC), Lotte van der Jagt (CCARE), Gregor Jurak (ULJ), Maroje Sorić (ULJ), Dominici Gabriele (USI), Renato Sabbadini (VUB), Marialetizia Mari (TRUST-IT), Anna Vilanova (TUE), Tuomas Kukko (JAMK), Teatske Altenburg (VUMC), Tanja Vrijkotte (VUMC (AMC)), Elena Mancuso (ENG) |
| Due Date | 2023.10.31 |
| Date | 2023.10.30 |
| Version | 1.0 DRAFT NOT YET APPROVED BY EUROPEAN COMMISSION |

**Dissemination Level**

| X | PU – Public, fully open |
|---|---|

| | 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 |

| **Abstract:** | This document provides information on the Data Management Plan of the project. It aims to provide the best practices and related standards for the data that is generated, processed, or collected in the context of SmartCHANGE, providing the principles towards the FAIR paradigm in accordance with the official EC guidelines. The document also reports security and ethics-related guidelines and processes that will be followed, emphasizing related regulations. |
|---|---|
| **Keyword List:** | Data management plan, FAIR, Findable, Accessible, Interoperable, Reusable, High-quality data, data security, ethics |
| **Licensing information:** | This work is licensed under Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)<br><br>http://creativecommons.org/licenses/by-sa/3.0/ |
| **Disclaimer:** | This project (GA No. 101080965) has received funding from the Horizon Europe R&I programme. The information provided in this document reflects solely the author's views. The European Community, Agency, and Commission are not liable for any use that may be made of the information contained herein. The content is provided without any guarantee or warranty of fitness for a particular purpose. Users utilise the information at their own risk and liability. In the case of proprietary information of the SmartCHANGE Consortium, it shall not be used, duplicated, or communicated to third parties without prior consent. |

# Versioning history

| Version | Date | Author(s) | Notes &/or Reason |
|---------|------|-----------|-------------------|
| **0.1** | 30/08/2023 | Thanos Kiourtis (UPRC), Argyro Mavrogiorgou (UPRC) | TOC and V0.1 |
| **0.2** | 30/09/2023 | Thanos Kiourtis (UPRC), Argyro Mavrogiorgou (UPRC) | Provided input to the related Sections of the document |
| **0.3** | 10/10/2023 | Anton Gradišek (JSI), José Carlos Ribeiro (UPORT), Thanos Kiourtis (UPRC), Argyro Mavrogiorgou (UPRC), Lotte van der Jagt (CCARE), Gregor Jurak (ULJ), Maroje Sorić (ULJ), Dominici Gabriele (USI), Renato Sabbadini (VUB), Marialetizia Mari (TRUST-IT), Anna Vilanova (TUE), Tuomas Kukko (JAMK), Teatske Altenburg (VUMC), Tanja Vrijkotte (VUMC (AMC)), Elena Mancuso (ENG) | Provided replies to the related questionnaire for the data that each organization is managing/collecting/processing |
| **0.4** | 19/10/2023 | Thanos Kiourtis (UPRC), Argyro Mavrogiorgou (UPRC) | Finalized deliverable and sent for internal review |
| **0.5** | 20/10/2023 | John Favaro (Trust-IT), Renato Sabbadini (VUB) | Internal review |
| **1.0** | 29/10/2023 | Thanos Kiourtis (UPRC), Argyro Mavrogiorgou (UPRC) | Updates based on internal review and finalization |

## Quality Control (peer & quality reviewing)

| Version | Date | Name (Organisation) | Role & Scope |
|---------|------|---------------------|--------------|
| **1.00** | 24/10/2023 | John Favaro (Trust-IT) | Reviewer |
| **1.00** | 27/10/2023 | Renato Sabbadini (VUB) | Reviewer |

# Table of contents

## List of tables

# Executive summary

*The current deliverable (D1.2) defines and describes the overall Data Management Plan (DMP) of SmartCHANGE. The DMP seeks to identify the best practices and specific standards for the generated data and assess their suitability for sharing and reuse in accordance with official guidelines. To this end, it aims to support the data management lifecycle for all data that will be collected, processed, or generated by the project to maximize its access.*

*SmartCHANGE's DMP will comply with the European Commission's (EC) DMP template [HE DMP, 2023], as it was refined for Horizon Europe, and will specify how the generated data will be easily discovered and accessed, ensuring open access and interoperability. It reflects the status of the data that is collected, processed, or generated, their respective data methodology and standards, whether and how these data will be shared and made open and how they will be curated and preserved.*

*This is the first version of SmartCHANGE's DMP produced, with a concrete final version becoming available on the third year of the project (M36).*

# List of abbreviations

| Abbreviation | Definition |
| --- | --- |
| AI | Artificial Intelligence |
| AIA | Artificial Intelligence Act |
| ALTAI | Assessment List for Trustworthy Artificial Intelligence |
| API | Application Programming Interfaces |
| CA | Consortium Agreement |
| CC | Creative Commons |
| CC0 | Creative Commons No Rights Reserved |
| CC-BY-NC-ND | CC attribution, non-commercial, non-derivative license |
| CESSDA | Consortium of European Social Science Data Archives |
| DMP | Data Management Plan |
| DOI | Digital Object Identifiers |
| DPIA | Data Protection Impact Assessment |
| DPO | Data Protection Officer |
| EB | Ethical Board |
| EC | European Commission |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| GA | Grant Agreement |
| GDPR | General Data Protection Regulation |
| HTML | HyperText Markup Language |
| IPR | Intellectual Property Rights |

| | |
|---|---|
| **JSON** | JavaScript Object Notation |
| **N/A** | Not Applicable |
| **PID** | Persistent Identifiers |
| **SELP** | Societal, Ethical, Legal and Privacy |
| **TFEU** | Treaty on the Functioning of the European Union |
| **TL** | Task Leader |
| **WP** | Work Package |
| **WPL** | Work Package Leader |
| **XML** | Extensible Markup Language |

# 1   Introduction

In European Union (EU) initiatives, a Data Management Plan (DMP) is created to outline how data will be gathered, processed, and/or generated, as well as how it will be kept. As it methodically outlines how data will be used by project partners and what data will be generated within the project's parameters, it is a crucial part of efficient data management. This will facilitate the project's seamless functioning while also motivating the project partners to make data management decisions based on SmartCHANGE and Findable, Accessible, Interoperable, and Reusable (FAIR) principles [Go FAIR, 2023].

SmartCHANGE aims at producing Artificial Intelligence (AI)-based long-term health risk evaluation for driving behaviour change strategies in children and youth, and in this context the current deliverable focuses on the management of the data collected, processed, or generated within the project and its lifecycle by all its consortium partners. The DMP aims to monitor the generated data in terms of their privacy and confidentiality, making sure that the legal and ethical standards for data generation, use, storage, and sharing are met throughout the project lifecycle and in accordance with the overall management of SmartCHANGE, as also provided in the Grant Agreement (GA) and Consortium Agreement (CA).

It should be made clear that when speaking about data in a DMP, this does not refer solely to datasets. More specifically, it may refer to:

- Internal administrative data, such as: (i) Data from administrative and financial management (e.g., partner contact information), (ii) Meeting/Web conference related material (e.g., participant's list, agenda, meeting minutes kept during a conference call or a general assembly meeting), (iii) Templates (e.g., deliverables' templates, presentations' templates), as well as (iv) Documentation of communication among members of the project (e-mails, instant messaging) (e.g., e-mails exchanged internally for SmartCHANGE communication purposes)
- External communication data, such as: (i) Deliverables and reports (e.g., SmartCHANGE deliverables), (ii) Publications (e.g., conference papers, journal articles), also including (iii) Data from the marketing/dissemination/communication/commercialization process (e.g., blog posts, social media posts, photographs from participation to events)
- Technical data, namely: (i) Software code (e.g., source code), and (ii) Prototype (e.g., solution prototypes)
- Proof-of-concept data, such as: (i) Datasets (e.g., open datasets, proprietary datasets, visualizations)

Information to produce this deliverable was gathered by the SmartCHANGE consortium through two (2) questionnaires (Appendix I, Appendix II) with a variety of questions per data

management aspect and in respect to the FAIR data principles. The structure of the DMP is also based on European Commission's (EC) DMP template. The DMP reflects a current picture of SmartCHANGE based on the answers from each partner and should be viewed as a living document that will be supplemented and refined as new issues and information come to light during the project's lifecycle.

## 1.1 Objective of the deliverable

SmartCHANGE's DMP aims to identify and organize all data collected, processed, or generated throughout the project's lifecycle, as well as specific standards used, data management policies and potential best practices showcased by the consortium partners.

## 1.2 Structure

The rest of the deliverable is structured as follows:

- Section 2 provides the methodology of this work and how the FAIR principles will be applied in the project, including basic terminologies that are used throughout the current deliverable.
- Section 3 provides an overview of the SmartCHANGE data, providing additional information on the types and content of the collected and generated data.
- Section 4 focuses on the FAIR principles in SmartCHANGE, including the techniques that will be used for making data findable, accessible, interoperable, and reusable.
- Sections 5 briefly replies to the questions related to the allocation of resources.
- Section 6 and Section 7 cover all the ethical and security aspects.
- Section 8 concludes the current deliverable, introducing our lessons learned as well as the next steps towards the 2nd version of the Data Management Plan (to be released in the 36th month of the project).

# 2 Methodology and basic terminology

## 2.1 FAIR management of research data

One difficulty that data-driven projects like SmartCHANGE face is making it easier for people to share expertise. To improve the FAIR paradigm of digital assets, the FAIR Principles for Scientific Data Management and Stewardship [Wilkinson, 2023] were released in 2016. In this part, the recommended practices for making data FAIR in SmartCHANGE are introduced. The FAIR principles, among other things, offer a continuum of growing reusability for data and digital objects across a variety of implementations (including data-related algorithms, tools, workflows, protocols, services, and other types of digital and research items). These solutions outline the capabilities and features of systems and services that can be used to produce beneficial research outputs, have them reviewed, and have them widely repurposed while giving the creators and users proper credit. Well-described, accessible, and standard data are essential for finding relevant data, performing machine analysis, and employing AI.

The FAIR principles have gained a lot of support from EU authorities as a quality standard and sparked discussions on data stewardship in open science and data-driven research around the world [EC Turning FAIR into reality, 2023]. They have moreover forced financing organizations to talk about how they will carry out their plans. Some of these needs are currently being established, but others already have a fully developed set of rules. This strategy typically comes before implementation decisions and does not imply a specific set of technical specifications, guidelines, or solutions. However, they place a focus on "machine actionability" which is the capacity of a computer system to locate, access, interact with, and reuse data without much help from humans. A summary of the FAIR guiding principles can be found in Table 1.

*Table 1 – FAIR guiding principles*

| FAIR feature | Principles |
|---|---|
| **Findable** | <ul><li>A globally unique and persistent identifier is given</li><li>Rich metadata are describing data</li><li>The identifier of the data that it represents is included in the metadata clearly and openly</li><li>(Meta)data is logged or indexed in a source that can be searched</li></ul> |
| **Accessible** | <ul><li>Using a standardized communication protocol, (meta)data can be retrieved by their identifier</li><li>The protocol is open, cost-free, and widely deployable</li></ul> |

| | |
|---|---|
| | • When necessary, the protocol enables for an authentication and authorization process |
| | • Metadata can still be accessed even if the underlying data is unavailable |
| **Interoperable** | • (Meta)data represents knowledge using a formal, understandable, shared, and broadly applicable language |
| | • (Meta)data employ FAIR-compliant vocabularies |
| | • (Meta)data contains appropriate references to other metadata |
| **Reusable** | • (Meta)data are accurately and pertinently described with a variety of rich descriptions |
| | • (Meta)data are made available under a transparent and easily understandable data usage license |
| | • Detailed provenance is linked to (meta)data |
| | • (Meta)data adhere to applicable community standards for the domain |

Simply expressed, the *Findability* principle dictates that data must be clearly and unambiguously identified, described, registered, or indexed. The primary properties of the data should be methodically documented, preferably in a standard format, and they should be indexed and stored in a public repository, such as a data archive. The data should also be given a distinctive and persistent identification. Data should be accessible through a clearly defined access mechanism, preferably through automated means, according to the principle of *Accessibility*. It entails setting up authentication and authorisation processes for access as well as, if necessary, automated protocols for data retrieval. Data and metadata must be conceptualized, communicated, and organised in accordance with widely accepted, publicly accessible standards based on the *Interoperability* principle. Standard data formats, variables, ontologies, and other techniques shall be used. The idea of *Reusability* states that data features and provenance should be thoroughly specified in accordance with community standards relevant to the domain, and that usage instructions should be simple to understand. Accurate and pertinent data descriptions, access and use permissions, community standards, and provenance for each data item are all required as part of this process.

Data life cycle management should use these four (4) principles, which are closely interrelated. The open sharing of data is not always required when the FAIR principles are applied to the workflow of a research effort. In this context, it should be made clear that open data and FAIR data are not the same thing. Open data is "*information that is freely available for anyone to access, use, and share for any purpose*". However, the FAIR principles offer "*recommendations for sharing data while also respecting any potential legal, moral, and contractual limitations*" [Jacobsen, 2020].

## 2.2　High-quality data generation best practices

The following best practices are intended to assist project partners in generating high-quality data, which will enhance research production and impact. It will give the data more structure so that other researchers can utilize it again. Additionally, it will enhance peer review and provide value to research, maintaining high levels of research integrity. Applying the FAIR data principles will therefore increase SmartCHANGE's transparency and reproducibility. FAIR principles are applied differently depending on the discipline and research method. However, there are several common elements to consider within search workflows such as:

- **Documentation**: By giving context, good documentation will guarantee that the data are understood by others.
- **Metadata**: When data is accompanied by metadata, it is easier to find it.
- **Data formats**: When integrating data or converting data to a machine-readable format, several formats must be taken into account.
- **Access to data**: The user controls who has access to which data and under what circumstances.
- **Persistent identifiers**: Through the creation of a link that remains with the data for a long time, persistent identifiers make it possible for others to locate and use the data.
- **Data licenses**: The legal contract defining how other parties may use the data is referred to as a license.

All in all, persistent identifiers and metadata are required to make data *Findable*. For data to be *Accessible*, it is important to specify who has access to the data and how. If it is not possible to publish data publicly due to data protection and IPR constraints, it is advised to enable access to metadata, for instance through a data repository. To promote data *Interoperability*, common standards and open data formats should be employed. Finally, the data can be *Reused* in a variety of ways and will be easier for others to understand if suitable documentation is added to them.

The latter is also in line with Directive (EU) 2019/1024 [Gobbato, 2020] on open data and re-use of public sector information (Open Data Directive) and Implementing Regulation (EU) 2023/138 [Florio, 2023] laying down a list of specific high-value data and their arrangements for their publication and re-use and in particular, the applicable conditions for re-use and the minimum requirements for disseminating data via Application Programming Interfaces (APIs).

Following, a description of the best practices for generating data of high-quality is being provided.

### 2.2.1  Documentation

Data documentation for research might be difficult. Therefore, it is critical to agree on a method for organizing and recording the data with input from all the project participants, to retain an accurate record of the data's generation or collection, to use a registration sheet, and to provide each object a unique identity. This could take the shape of techniques, tools, or software. Partners in the project should make sure to record the entire "research process" in writing. For instance, it would be helpful to know who handled the data, how it was utilized, what the study's conclusion was, how these data relate to other data, or what papers have been released. A thoroughly documented research process gives the data context, making it easier for future researchers to interpret and comprehend it.

### 2.2.2  Metadata

Another important type of documentation is metadata, or information that characterizes other information. Metadata includes statements on the descriptive, contextual, and provenance nature of research data or digital artifacts [Ulrich, 2022]. The location, date, time, file size, file format, keywords, and technical characteristics are just a few examples. For data to be accessible, comprehended, and processed by both machines and people, metadata must be present. For instance, an algorithm identifies text and associates it with other information that has a similar theme. Metadata has long been a concern for researchers, and several academic fields have created standards for meta descriptions. Metadata standards work to provide a shared understanding of the meaning of the data to guarantee the correct and accurate use and interpretation of the data.

In the case of SmartCHANGE, metadata standards are necessary to perform technological research since consistency and the capacity for automatic data processing can both be improved using metadata. In most data combinations, most of the effort spent preparing the data is spent interpreting it. As a result, using standard metadata will cut down on the time needed for data preparation. Furthermore, well-designed metadata will ensure that data are easily findable, interoperable, and reusable across systems, disciplines, and languages. Within the SmartCHANGE research domain, metadata standards should be applied and should be based on best practices or those that are generally acknowledged and utilized. Project partners should include all the information necessary for comprehending the data in the absence of any common standards and automatic metadata production. Think about creating a "Readme" file, for instance, which includes metadata describing the data in a way that is likely to be understood by others.

### 2.2.3   Data formats

Data formats are helpful for gathering and exchanging data with other academics. Data can occasionally be provided in an odd binary format. Data interoperability is improved and reused by researchers and machines, and it is made simpler when data is stored in generally used formats. As a result of the FAIR principles, the reuse of research data is maximized.

### 2.2.4   Access to data

If there are any limitations due to data protection or Intellectual Property Rights (IPR), for example, the metadata should be accessible. If the data can be accessed, how can they be accessed, and who can be contacted for access requests? That information should be included in the metadata.

### 2.2.5   Persistent identifiers

To ensure that data are easily accessible and findable, a persistent identifier for the data should be assigned to it. As an example, Digital Object Identifiers (DOIs), which are frequently used when citing works in scholarly journals, point researchers to the main website where an article has been published, giving them access if the material was made publicly available.

### 2.2.6   Data licenses

A license determining how the data may be used by others should be provided when making data openly available, with several types of data licenses being available (e.g., Creative Commons (CC)). Depending on the license, researchers can either share data openly or restrictively. For instance, a CC license follows the "some rights reserved" model in which, the copyright holder may retain a few rights from the bundle conferred by law. The fact that researchers can share (copy and redistribute) the content in any format or medium makes the CC BY 4.0 license perhaps the most appealing of the several CC licenses. Additionally, they could alter things for free or even for profit. But they have to say if changes were made, include a link to the license, and give due credit to the inventor. There are additional data licenses, such as Creative Commons No Rights Reserved (CC0), or the CC attribution, non-commercial, non-derivative license (CC-BY-NC-ND), that provide different types of limitations on the use of data.

## 2.3   Collection of information for SmartCHANGE DMP

The information for this deliverable was gathered by the SmartCHANGE consortium through a questionnaire dedicated to all the SmartCHANGE partners (Appendix I), as well as a

questionnaire dedicated to the pilot partners (UPORTO, VUMC, JAMK, ULJ,) in SmartCHANGE (Appendix II). Both are based on the "Guidelines on FAIR Data Management" [Wilkinson, 2016] and corresponds to an updated template associated with them (Horizon Europe Data Management Plan Template), with some additional information requested. The leaders of this deliverable gave the partners the final version of the questionnaires after making several changes. Although SmartCHANGE's partners answered quickly to the questionnaires, not all the needed details were available or sufficiently detailed at this early stage of the project. Being a live document, all modifications are logged throughout the project and will be incorporated into the DMP's upcoming version.

# 3    SmartCHANGE data summary

The following Sections provide information on the data that has been collected and generated based on the project's activities.

## 3.1    Types of collected and generated data

In SmartCHANGE both quantitative (i.e., data in the form of numbers) and qualitative (i.e., non-numerical) data are being collected. The identified data can be classified according to the following typology:

- Data collected/generated through internal administration procedures.
- Data collected/generated for external communication procedures.
- Data collected/generated for technical purposes.
- Data collected/generated for proof-of-concept purposes.
- Data collected/generated for SmartCHANGE evaluation.

Table 2 illustrates data that will be generated by SmartCHANGE during the first years of the project's lifetime. This list is subject to several updates till the end of the project.

*Table 2 – Data to be collected and/or generated in SmartCHANGE*

| Data name | SmartCHANGE partner |
|---|---|
| *Data collected/generated through internal administration procedures* | |
| Internal communication data | ALL |
| Partner's data | ALL |
| Internal administrative data | ALL |
| **Data collected/generated for external communication procedures** | |
| Deliverables | ALL |
| Publications | ALL |
| Users' registration data | ALL (when appropriate) |
| **Data collected/generated for technical purposes** | |
| Solution Prototypes | Technical Partners |
| Software code | Technical Partners |
| Co-design activities data | ALL |
| **Data collected/generated for proof-of-concept purposes** | |
| Healthcare/ Proof-of-concept data | ULJ, VUMC, VUMC (AMC), UPORTO, JAMK |
| **Data collected/generated for SmartCHANGE evaluation** | |
| SmartCHANGE Evaluation/Recommendation data | ALL |

## 3.2    Formats of the data

The collected and generated data will be produced in various formats, depending on the nature and purpose of the data collected. Wherever possible, SmartCHANGE intends to give preference to open, non-proprietary standards (including de facto standards), hence enhancing the data's interoperability and potential for reuse. The forms and formats in which the various SmartCHANGE data can be found are shown in the following table (Table 3).

*Table 3 – Data formats of the SmartCHANGE data*

| Data name | Format |
| --- | --- |
| *Data collected/generated through internal administration procedures* | |
| Internal communication data | .msg, .pptx, .docx, .xlsx, .pdf |
| Partner's data | .docx, .xlsx |
| Internal administrative data | .docx, .pdf |
| *Data collected/generated for external communication procedures* | |
| Deliverables | .docx, .pdf |
| Publications | .docx, .pdf, .latex |
| Users' registration data | Drupal database format |
| *Data collected/generated for technical purposes* | |
| Solution prototypes | .pptx, Adobe XD file format |
| Software code | .ts, .tsx, .mts, .cts, .java, .py, .ipynb, .c++ |
| Co-design activities data | .jpeg, .pdf and formats derived from dedicated questionnaires, interviews, mock-up testing, trade-off experiments |
| *Data collected/generated for proof-of-concept purposes* | |
| Healthcare/ Proof-of-concept data | .spss, .xlsx, .csv, .sav, .txt, .dat |
| *Data collected/generated for SmartCHANGE evaluation* | |
| SmartCHANGE Evaluation/Recommendation data | .docx, .pdf, .mp3 |

## 3.3    Origin of data and re-use of pre-existing data

In SmartCHANGE, the new data are being collected/generated in the following ways:

- By the consortium partners (desk research and respective analysis and assessments)
- Through contacts with outside parties taking part in the project's activities, as well as the Advisory Board, Experts, and Interest Groups. External stakeholders from whom new data will originate include the following groups (Table 4):

*Table 4 – External stakeholders' groups*

| Stakeholders group | Description |
|---|---|
| **Researchers and Academy** | Researchers from industry and academia |
| **Healthcare funded projects** | Projects and organisations involved in these past and ongoing initiatives |
| **SMEs/Startups and IoT Innovators** | Public and private organisations and Startups/SMEs, innovators |
| **Standardisation bodies, pre-standardisation and Open-source initiatives** | Standardisation organisations, working groups, pre-standardisation groups |
| **Security and privacy stakeholders** | Working groups, researchers, Technology providers, IoT infrastructure providers |
| **EU Initiatives from relevant domains/ technologies** | Healthcare and related domains initiatives, Initiatives aggregating innovators |
| **Policymakers, regulators and public authorities** | Actors committed to support the development of the full economic potential of Europe |

## 3.4   Size and Volume of the data

The size of each data can be found in Section 3.6, for each different data type.

## 3.5   Data utility

All the project's stakeholders may find one or more of the data collected or produced during SmartCHANGE to be of particular value and may gain benefit from having access to and reusing this data. The tables created for each data (Table 5 – Table 23) that is anticipated to be collected or generated in the context of the project in the following Section (Section 3.6) provide more specific information on the stakeholder groups who may be interested in the reuse of each data.

## 3.6   SmartCHANGE data

### 3.6.1   Data collected/generated through internal administration procedures

*Table 5 – Internal communication data*

| Internal communication data | Description |
|---|---|
| Data description | This data is related with documentation of communication among members of the consortium (e.g., e-mails exchanged internally for SmartCHANGE communication purposes, Meeting/Web conference related material) |

| Related WP/Task | ALL |
|---|---|
| Data origin | Communication among consortium partners |
| Will you re-use any existing data? If yes, how? | Materials produced in other parts of the project will be reused in the construction of this data |
| Methodologies for data collection/generation | Collaboration among project partners |
| Data format | The data will be provided in: .msg, .pptx, .docx, .xlsx, .pdf |
| Where this data will be stored? | Data will be stored internally by the project team in the password-protected document repository. Key outputs of these contributions are shared in the context of public deliverables or as events' reports |
| Expected size of the data | Order of Gigabytes |
| Metadata and standards used | Each data source is accompanied by the following minimum set of metadata: (i) Title, (ii) Author, (iii) Creation date and purpose, and (iv) Usage - relevant group and meeting date |
| For whom might the data be useful? | This data will be useful for SmartCHANGE partners |
| Data access, sharing and licensing | The mailing list conversations and contacts of participants will remain confidential as required under GDPR |

*Table 6 – Partner's data*

| Partner's data | Description |
|---|---|
| Data description | This data is related with information and knowledge that will be gathered from questionnaires that will be shared to the consortium partners. |
| Related WP/Task | WP2, WP4, WP5 |
| Data origin | Replies to specific questionnaires |
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | Organization of webinars, and online training that will lead the process of collecting data through questionnaires |
| Data format | The data will be provided in: .docx, .xlsx |
| Where this data will be stored? | Data will be stored internally by the project team in the password-protected document repository. Key outputs of these contributions are shared in the context of public deliverables or as events' reports |
| Expected size of the data | Order of Megabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for SmartCHANGE partners |
| Data access, sharing and licensing | N/A |

*Table 7 – Internal administrative data*

| Internal administrative data | Description |
|---|---|
| Data description | This data is related with administrative data that is collected internally among the consortium partners for record keeping and possibly statistical analysis |
| Related WP/Task | ALL |
| Data origin | Administrative information collected from partners |
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | Collaboration among project partners |
| Data format | The data will be provided in: .docx, .xlsx, .pdf |
| Where this data will be stored? | Data will be stored internally by the project team in the password-protected document repository |
| Expected size of the data | Order of Megabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for SmartCHANGE partners |
| Data access, sharing and licensing | Internal, restrictive access available on case-by-case basis upon request |

### 3.6.2 Data collected/generated for external communication procedures

*Table 8 – Deliverables*

| Deliverables | Description |
|---|---|
| Data description | This data is related with a report/document created as a result of SmartCHANGE and intended for delivery to the EC |
| Related WP/Task | ALL |
| Data origin | Collaboration among SmartCHANGE partners |
| Will you re-use any existing data? If yes, how? | Data originated from previous/on-going project activities of the involved partners in the form of baseline knowledge |
| Methodologies for data collection/generation | Collaboration among project partners in which each responsible for the deliverable (Task leader) will assign dedicated tasks to the involved partners |
| Data format | The data will be provided in: .docx, .pdf |
| Where this data will be stored? | Data will be stored internally by the project team in the password-protected document repository. It will be also shared in the https://www.smart-change.eu/ and Zenodo platform |
| Expected size of the data | Order of Gigabytes |
| Metadata and standards used | Each data source is accompanied by the following minimum set of metadata: (i) Title, (ii) Author, (iii) Creation date and purpose, and (iv) Usage - relevant group and meeting date |

| | |
|---|---|
| For whom might the data be useful? | This data will be useful for SmartCHANGE external stakeholders and project partners |
| Data access, sharing and licensing | This data will become freely / openly accessible based on an open license. In the case that a deliverable will be marked as confidential, it will be only accessible from the project partners |

*Table 9 – Publications*

| *Publications* | *Description* |
|---|---|
| Data description | This data is related with a report/document intended to further the progress of science, usually by sharing findings from research with readers |
| Related WP/Task | ALL |
| Data origin | Collaboration among SmartCHANGE partners |
| Will you re-use any existing data? If yes, how? | Data originated from previous/on-going project activities of the involved partners in the form of baseline knowledge |
| Methodologies for data collection/generation | Collaboration among project partners in which the involved publication authors will provide their research expertise into compiling the final document |
| Data format | The data will be provided in: .docx, .pdf, .latex |
| Where this data will be stored? | Data will be stored internally by the project team in the password-protected document repository. It will be also indexed in publishers' websites (e.g., Elsevier, Springer, IEEE) whereas it will be also shared in the https://www.smart-change.eu/ and Zenodo platform. In some cases, if it is appropriate, it will be shared in Open Access repositories (e.g., Institutional repositories, MedRxiv, European Open Science Cloud) |
| Expected size of the data | Order of Megabytes |
| Metadata and standards used | Each data source will be accompanied by the sets of metadata that are required from the different publishers |
| For whom might the data be useful? | This data will be useful for SmartCHANGE external stakeholders (e.g., scientific community, policymakers, health professionals working with children) and project partners |
| Data access, sharing and licensing | This data will be accessible as it will be required from the different publishers' policies |

*Table 10 – Users' registration data*

| *Users' registration data* | *Description* |
|---|---|
| Data description | This data is related with personal data that are utilized to identify a registered user to the SmartCHANGE website and related newsletter |
| Related WP/Task | WP8 |

| Data origin | Data collected from https://www.smart-change.eu/user/register (user registration) and https://www.smart-change.eu/ (newsletter subscription) |
|---|---|
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | Registration form compiled by the user |
| Data format | The data will be provided in: Drupal database format |
| Where this data will be stored? | Data will be stored in the Drupal database (MySQL) on AWS Eu-west-1 |
| Expected size of the data | Order of Megabytes (less than 1 Megabyte per registered user) |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for the SmartCHANGE consortium for the scope listed in the privacy policy statement (https://www.smart-change.eu/privacy-policy) |
| Data access, sharing and licensing | Data will be accessed by the registered user, by the Data controller and by the Data Processor identified in the Privacy Policy (https://www.smart-change.eu/privacy-policy) |

### 3.6.3 Data collected/generated for technical purposes

*Table 11 – Solution prototypes*

| Solution prototypes | Description |
|---|---|
| Data description | This data is related with an early sample, model, or release of a product or system built to test a concept or assumption or to act as a thing to be tested and learned from |
| Related WP/Task | WP4, WP5, and WP6 |
| Data origin | Collaboration among SmartCHANGE partners |
| Will you re-use any existing data? If yes, how? | Data originated from previous/on-going project activities of the involved partners in the form of baseline knowledge |
| Methodologies for data collection/generation | Collaboration among project partners in which each technical partner will provide his technical knowhow and background expertise |
| Data format | The data will be provided in: .pptx, Adobe XD file format |
| Where this data will be stored? | Data will be stored internally by the project team in a password-protected code repository – this will be determined as the project will become more mature |
| Expected size of the data | Order of Gigabytes |
| Metadata and standards used | General code quality standards |
| For whom might the data be useful? | This data will be useful for SmartCHANGE external stakeholders (e.g., scientific community, policymakers, health professionals working with children) and technical project partners |

| Data access, sharing and licensing | This data will become freely/openly accessible based on an open license by the time it will be considered appropriate |

*Table 12 – Software code*

| Software code | Description |
|---|---|
| Data description | This data is related with any collection of text, with or without comments, written using a human-readable programming language, usually as plain text |
| Related WP/Task | WP4, WP5, and WP6 |
| Data origin | Collaboration among SmartCHANGE partners |
| Will you re-use any existing data? If yes, how? | Data originated from previous/on-going project activities of the involved partners in the form of baseline knowledge |
| Methodologies for data collection/generation | Collaboration among project partners in which each technical partner will provide his technical knowhow and background expertise |
| Data format | The data will be provided in: .ts, .tsx, .mts, .cts, .java, .py, .ipynb, .c++ |
| Where this data will be stored? | Data will be stored internally by the project team in a password-protected code repository – this will be determined as the project will become more mature |
| Expected size of the data | Order of Gigabytes |
| Metadata and standards used | General code quality standards |
| For whom might the data be useful? | This data will be useful for SmartCHANGE external stakeholders (e.g., scientific community, policymakers, health professionals working with children) and technical project partners |
| Data access, sharing and licensing | This data will become freely/openly accessible based on an open license by the time it will be considered appropriate |

*Table 13 – Co-design activities data*

| Co-design activities data | Description |
|---|---|
| Data description | This data is related with data collected during co-design sessions with parents and their children, as well as collected during co-design sessions with health professionals |
| Related WP/Task | WP3 |
| Data origin | Co-design sessions among SmartCHANGE related partners and parents and their children/health professionals |
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | Co-design sessions according to protocol described in SmartCHANGE deliverable D3.1 (Participatory Design Protocol) |
| Data format | The data will be provided in: .jpeg, .pdf |

| Where this data will be stored? | Physically generated data (e.g., drawings, text) will be transformed to digital form and stored in the password-protected project's repository. |
|---|---|
| Expected size of the data | Order of Gigabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for designers of the SmartCHANGE application for families, AI partners working on explainability of the AI models, as well as the mHealth industry |
| Data access, sharing and licensing | This data will be accessible only to dedicated partners from the consortium involved in the co-design process |

### 3.6.4    Data collected/generated for proof-of-concept purposes

#### 3.6.4.1    UPORTO: Proof-of-concept data

*Table 14 – UPORTO: Proof-of-concept data*

| *UPORTO: Proof-of-Concept data* | *Description* |
|---|---|
| Data description | This data is related to datasets from children and adolescents, that were collected in several research projects over the last years, and from FitEscola |
| Related WP/Task | WP4, WP7 |
| Data origin | The origin of the data is mainly data collection in school settings |
| Will you re-use any existing data? If yes, how? | These datasets were obtained a few years ago during some research projects, and there is no plan to reuse them. However, since UPORTO is responsible for datasets merging, the partners are still not sure whether they will only get the information about the datasets or not, and whether the data itself will be sent to other partners or it will be sent to them instead |
| Methodologies for data collection/generation | Objective and subjective measures were used for data collection, that includes questionnaires, physical evaluations, and blood samples for metabolic risk profiles |
| Data format | The data will be provided in: .spss, .csv |
| Where this data will be stored? | Currently the data is stored on UPORTO servers. However, there is not for the moment any update about where the data will be stored for SmartCHANGE purposes |
| Expected size of the data | Order of Gigabytes |
| Metadata and standards used | The metadata includes a codebook of the datasets. |
| For whom might the data be useful? | This data will be useful for SmartCHANGE external stakeholders (e.g., scientific community, policymakers, health professionals working with children) and technical project partners |

| Data access, sharing and licensing | All the data will be available only after pseudo-anonymization of the participants. It will be available only at the password-protected project repository to the members of the consortium who are directly involved in data processing, and only for specific aims of the project |
|---|---|
| **Additional information** | |
| Is the data personal? | Yes, but it will be pseudo-nymized |
| Is the data sensitive? | Yes, although not trackable without a codebook only in UPORTO's possession, and not shareable with other partners for privacy purposes |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | Susana Pereira<br>dados.pessoais@up.pt |
| Is there any data protection policy assigned to this data? If yes, provide a copy | N/A |

### 3.6.4.2    VUMC (AMC): Proof-of-concept data

*Table 15 – VUMC (AMC): Proof-of-concept data*

| VUMC (AMC): Proof-of-Concept data | Description |
|---|---|
| Data description | Datasets from the ABCD study, that will be used to build the AI model for SmartCHANGE |
| Related WP/Task | WP4, WP7 |
| Data origin | The data has been collected among ABCD participants |
| Will you re-use any existing data? If yes, how? | The data has been collected a few years ago by VUMC (AMC) |
| Methodologies for data collection/generation | The datasets have been collected through questionnaires and physical measurements |
| Data format | The data will be provided in: .spss |
| Where this data will be stored? | The original datasets will be stored on the VUMC (AMC) servers. The partners do not know yet where it will be stored for SmartCHANGE. VUMC (AMC)'s ICT and legal departments do not allow storage of study data on the password-protected project's repository; therefore, this would not be an option. |
| Expected size of the data | This is not known yet |
| Metadata and standards used | The metadata includes a codebook of the datasets. No standards are used |
| For whom might the data be useful? | This data will be useful for technical project partners |
| Data access, sharing and licensing | These datasets will only be shared with partners with whom VUMC (AMC) will have signed an agreement - the datasets cannot be shared among partners |

| Additional information | |
|---|---|
| Is the data personal? | Yes |
| Is the data sensitive? | Yes |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | The ABCD-study is registered by the DPO, but it is not a common practice here to inform him of every data transfer. VUMC (AMC)'s legal department drafts and checks all the agreement and they will let VUMC (AMC) know if it is necessary to have it checked by the DPO |
| Is there any data protection policy assigned to this data? If yes, provide a copy | The privacy statement can be found here: Privacy statement of Amsterdam UMC |

### 3.6.4.3 JAMK: Proof-of-concept data

*Table 16 – JAMK: Proof-of-concept data*

| JAMK: Proof-of-Concept data | Description |
|---|---|
| Data description | Data from young Finns Study: a follow-up data from 1980 to 2018-2020 of health and behavioural variables of six cohorts (original N=3,596) |
| Related WP/Task | WP4, WP7 |
| Data origin | https://youngfinnsstudy.utu.fi/ |
| Will you re-use any existing data? If yes, how? | The data has been collected a few years ago from 1980 to 2018-2020 from young Finns |
| Methodologies for data collection/generation | Objective and subjective measures were used for data collection, which includes questionnaires, physical evaluations, and blood samples for metabolic risk profiles |
| Data format | The data will be provided in: .spss, .xlsx, .csv, .txt, .dat |
| Where this data will be stored? | For the moment the datasets are stored in JAMK servers, but it will be defined where all the data will be stored for SmartChange purposes |
| Expected size of the data | Order of Megabytes |
| Metadata and standards used | Variable descriptions will be included as metadata |
| For whom might the data be useful? | This data will be useful for technical project partners |
| Data access, sharing and licensing | The data will be available only at the project repository to those members of the consortium who are directly involved in data processing, and only for specific aims of the project |
| Additional information | |
| Is the data personal? | Yes, but it will be pseudo-anonymized |
| Is the data sensitive? | Yes |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | Olli Raitakari (University of Turku, responsible researcher, project coordinator) olli.raitakari@utu.fi |

| Is there any data protection policy assigned to this data? If yes, provide a copy | University of Turku owns the data. The data transfer agreement is needed between Univ. of Turku and collaborators working with the data set. (Contact information for responsible researcher: https://youngfinnsstudy.utu.fi/contactus.html) |

### 3.6.4.4 ULJ: Proof-of-concept data

*Table 17 – ULJ: Proof-of-concept data (SLOfit students)*

| SLOfit students | Description |
|---|---|
| Data description | Longitudinal data of near-census data on physical fitness (PF) of Slovenian school children, collected each year by the law in all Slovenian schools in so called "Sport Educational Chart (ŠVK) information system". Part of the data is stored in MS SQL database (used by My SLOfit web app), the other part locally (on institution hard disks). More details are available here |
| Related WP/Task | WP4, WP7 |
| Data origin | ŠVK information system (see above) |
| Will you re-use any existing data? If yes, how? | Before original data can be used in the project, it must be exported in anonymous form in agreed format (e.g. SPSS) |
| Methodologies for data collection/generation | Data are collected by a standard protocol in line with written instructions (supervised) by PF teachers. Description of the tests: https://en.slofit.org/measurements/test-battery |
| Data format | The data will be provided in: .spss |
| Where this data will be stored? | The data will be stored in the project's password-protected repository |
| Expected size of the data | 500 Megabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | The data will be useful for partners involved in risk prediction modelling tasks |
| Data access, sharing and licensing | Data will be available only at the project repository to those members of the consortium who are directly involved in data processing |
| **Additional information** | |
| Is the data personal? | No, data will be anonymized |
| Is the data sensitive? | No, since data will be anonymized |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | Nina Komočar Urbanija dpo@uni-lj.si |
| Is there any data protection policy assigned to this data? If yes, provide a copy | Yes, described in DPIA (Data Protection Impact Assessment; in Slovene) and https://www.slofit.org/o-slofit/varstvo-osebnih-podatkov (in Slovene) |

*Table 18 – ULJ: Proof-of-concept data (SLOfit lifelong)*

| SLOfit lifelong | Description |
|---|---|
| Data description | Longitudinal data of adult Slovenian citizens who choose to participate in SLOfit adults PF measurements (see test battery) and also agreed to post their (historical) ŠVK data collected during their school years. Additionally, some data about 24-hour movement behaviour and health status are collected by questionnaires. |
| Related WP/Task | WP4, WP7 |
| Data origin | https://en.slofit.org/Measurements/SLOfit-Adults-Measurements |
| Will you re-use any existing data? If yes, how? | Before original data can be used in the project, it must be exported in anonymous form in agreed format (e.g. SPSS) |
| Methodologies for data collection/generation | Data are collected by a standard protocol in line with written instructions (supervised) by PE teachers and other sport/PF professionals. Measurement protocol is described here (in Slovene only) |
| Data format | The data will be provided in: .spss, .sav |
| Where this data will be stored? | The data will be stored in the project's password-protected repository |
| Expected size of the data | 10 Megabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | The data will be useful for partners involved in risk prediction modelling tasks |
| Data access, sharing and licensing | Data will be available only at the project repository to those members of the consortium who are directly involved in data processing |
| **Additional information** | |
| Is the data personal? | No – data will be anonymized |
| Is the data sensitive? | No, since data will be anonymized |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | Nina Komočar Urbanija dpo@uni-lj.si |
| Is there any data protection policy assigned to this data? If yes, provide a copy | Yes, described in DPIA (Data Protection Impact Assessment; in Slovene) and https://www.slofit.org/o-slofit/varstvo-osebnih-podatkov (in Slovene) |

*Table 19 – ULJ: Proof-of-concept data (SLOfit students' lifestyle)*

| SLOfit students' lifestyle | Description |
|---|---|
| Data description | Longitudinal data on fitness and 24-hour movement behaviour of schoolchildren whose parents choose to register in My SLOfit web application. |
| Related WP/Task | WP4, WP7 |

| Data origin | The data originate from the My SLOfit web application |
|---|---|
| Will you re-use any existing data? If yes, how? | Before original data can be used in the project, it must be exported in anonymous form in agreed format (e.g., SPSS) |
| Methodologies for data collection/generatio | Data are collected within My SLOfit web application via questionnaire |
| Data format | The data will be provided in: .spss, .sav |
| Where this data will be stored? | The data will be stored in the project's password-protected repository |
| Expected size of the data | 1 Megabyte |
| Metadata and standards used | N/A |
| For whom might the data be useful? | The data will be useful for partners involved in risk prediction modelling tasks |
| Data access, sharing and licensing | Data will be available only at the project repository to those members of the consortium who are directly involved in data processing |
| **Additional information** | |
| Is the data personal? | No – data will be anonymized |
| Is the data sensitive? | No, since data will be anonymized |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | Nina Komočar Urbanija dpo@uni-lj.si |
| Is there any data protection policy assigned to this data? If yes, provide a copy | Yes, described in DPIA (Data Protection Impact Assessment; in Slovene) and https://www.slofit.org/o-slofit/varstvo-osebnih-podatkov (in Slovene) |

*Table 20 – ULJ: Proof-of-concept data (ACD.Si)*

| **ACD.Si** | **Description** |
|---|---|
| Data description | PF and lifestyle and health-related data of Slovenian school children 6-19-years-old, collected every 10 years on clustered sample (at the school level). More details are available here. |
| Related WP/Task | WP4, WP7 |
| Data origin | More details can be found here. |
| Will you re-use any existing data? If yes, how? | Before original data can be used in the project, it must be exported in anonymous form in agreed format (e.g. SPSS) |
| Methodologies for data collection/generation | Data are collected by a standard protocol (see here) |
| Data format | The data will be provided in: .spss, .sav |
| Where this data will be stored? | The data will be stored in the project's password-protected repository |
| Expected size of the data | 50 Megabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | The data will be useful for partners involved in risk prediction modelling tasks |

| Data access, sharing and licensing | Data will be available only at the project repository to those members of the consortium who are directly involved in data processing |
|---|---|
| **Additional information** | |
| Is the data personal? | No – data will be anonymized |
| Is the data sensitive? | No, since data will be anonymized |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | Nina Komočar Urbanija<br>dpo@uni-lj.si |
| Is there any data protection policy assigned to this data? If yes, provide a copy | Yes, described in DPIA (Data Protection Impact Assessment; in Slovene) and https://www.slofit.org/o-slofit/varstvo-osebnih-podatkov (in Slovene) |

### 3.6.5   Data collected/generated for SmartCHANGE recommendation/evaluation

*Table 21 – Data for the design of SmartCHANGE applications*

| *Data for the design of SmartCHANGE applications* | *Description* |
|---|---|
| Data description | Data on needs of users (adolescents, health care practitioners) to inform the design of the SmartCHANGE applications; according to co-creation protocol (D3.1) |
| Related WP/Task | WP3 |
| Data origin | The data will be collected from co-creation meetings with children and health care practitioners |
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | The methodologies that will be described in the co-creation protocol (D3.1) |
| Data format | The data will be provided in: .docx, .pdf, .mp3 |
| Where this data will be stored? | For physical data, all data will be locked in a cupboard on the VUMC premises. Digital or electronic data will be stored on password-protected computers and secure servers hosted by the VUMC |
| Expected size of the data | 10 Gigabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for technical project partners |
| Data access, sharing and licensing | These datasets will be only accessible to the relevant researchers on the project |

*Table 22 – Evaluation of SmartCHANGE applications*

| Evaluation of SmartCHANGE applications | Description |
|---|---|
| Data description | Data for the evaluation of the proof-of-concept study of the SmartCHANGE applications, including measures of feasibility (i.e., acceptability, demand, implementation, practicality, adaptation, integration, and expansion), user experience and usability, as well as explainability of the SmartCHANGE models. |
| Related WP/Task | WP7 |
| Data origin | The data will be collected during the co-creation meetings with the focus groups of the pilot sites |
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | Data will be collected via surveys, focus groups and interviews with different users of the ecosystem at the four sites (i.e., healthcare personnel, youth, parents) |
| Data format | The data will be provided in: .docx, .pdf, .mp3 |
| Where this data will be stored? | The data will be on the VUMC premises. Digital or electronic data will be stored on password-protected computers and secure servers hosted by the VUMC |
| Expected size of the data | 10 Gigabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for all the project partners |
| Data access, sharing and licensing | This has not been decided yet, since for the proof-of-concept studies, data deriving from different countries might need to be shared |

*Table 23 – Recommendations for the design of SmartCHANGE applications*

| Recommendations for the design of SmartCHANGE applications | Description |
|---|---|
| Data description | Data for the development of recommendations for a wider implementation of the SmartCHANGE tools for personalised prevention strategies |
| Related WP/Task | WP8 |
| Data origin | The data will be collected from focus group interviews with stakeholders involved in this proof-of-concept study in the four pilot countries (WP7), including health professionals, children/young people, and their parents. It will also include data from a Delphi study to collect input from a broader stakeholder group, with health professionals, policy makers and implementation experts from relevant national and international organizations and networks. |

| | |
|---|---|
| Will you re-use any existing data? If yes, how? | N/A |
| Methodologies for data collection/generation | It will be used both the Delphi study approach, as well interviews |
| Data format | The data will be provided in: .docx, .pdf, .mp3 |
| Where this data will be stored? | The data will be on the VUMC premises. Digital or electronic data will be stored on password-protected computers and secure servers hosted by the VUMC |
| Expected size of the data | 10 Gigabytes |
| Metadata and standards used | N/A |
| For whom might the data be useful? | This data will be useful for all the project partners |
| Data access, sharing and licensing | These datasets will be only accessible to the relevant researchers on the project |

# 4 FAIR data

To make research data FAIR, and to assure their sound administration, the projects funded under Horizon Europe are given guidance in the "Guidelines on FAIR Data Management" of the EC. This is necessary to enable the sharing of scientific information, knowledge, and the creation of new knowledge. The EU General Data Protection Regulation (GDPR)'s guarantees and protections will be followed while processing and making the data public [Johnson, 2022]. Prior to their anonymization and aggregation, all personally identifiable information gathered or created will be treated as closed data, protecting the privacy of the data subjects. Before deciding whether to make a particular set of data public or not, it will be evaluated for sensitivity, privacy, and security concerns. The methodology used by SmartCHANGE to make data Findable, Accessible, Interoperable, and allowing the greatest amount of Reuse is described in the following Sections.

## 4.1 Making data findable

Enhancing the *Findability* of the data gathered/generated throughout its operations is a priority for SmartCHANGE. The project makes use of a metadata-driven method to improve data searchability. The project's data collection and generation will be made available using metadata that is appropriate for the data's format and content.

### 4.1.1 Public data repository

SmartCHANGE will make use of Zenodo platform [Zenodo, 2023] to further strengthen the discoverability of the data designated as open. Zenodo is a free, public data repository that was developed by OpenAIRE and CERN under contract with the EC. The Zenodo service enables researchers, scientists, and EU initiatives to share their findings and promotes data reuse both within the EU and around the world. It can handle any file type with a maximum file size of 50 Gigabytes. By defining and preserving additional metadata provided by the uploader, Zenodo can improve the discoverability of the data collected and made public by SmartCHANGE. Additionally, it provides the option for the consortium to limit data access for a chosen user group (or the general public) for a predetermined amount of time. Through DataCite [DataCite, 2023], the foremost international non-profit organization providing Persistent Identifiers (PIDs) (and specifically DOIs) for research data and other research outputs, Zenodo registers and maintains DOIs for every submitted data. The foundation of the CERN data centre, which is secure and reliable, is used to maintain the submitted data. All these factors ensure that the data stored in Zenodo will be available in the future.

With this in mind, Zenodo's specified metadata standards will be applied to SmartCHANGE's publicly accessible data.

### 4.1.2   Naming conventions and versioning

To make the data gathered or generated by SmartCHANGE *Findable*, it is required to adhere to a consistent naming and versioning convention. Consistent naming and versioning practices improve data searchability, enabling interested parties like partners and stakeholders to quickly find the data they are looking for. The project's naming scheme for its data, in particular, can be summarized following the below example (or similar):

**HEU-101080965-SmartCHANGE_<data name>_<version>_<date>.<extension>**

- **<data name>**: short (3-4 words) data name
- **<version>**: the versioning number of a data
- **<date>**: dd.mm.yyyy
- **<extension>**: the extension of the corresponding data type

Every data file created throughout the course of the project will have a clear version number indicator in the naming convention to make it easier to distinguish between various versions.

### 4.1.3   Search keywords and metadata

To facilitate *Findability* of the SmartCHANGE data by interested parties, search keywords (i.e., words and phrases used to describe and interpret data and accurately reflect the content of the data) and metadata (i.e., structured textual information that describes something about the creation, content, or context of a digital resource) will be provided for the data.

## 4.2   Making data accessible

### 4.2.1   Data accessibility

Resources that are considered open are expected to be made available under a CC-BY v4.0 license, which permits community sharing of project resources as well as commercial reuse with proper credit. By encouraging people to use the materials created by SmartCHANGE as needed, this aids in their dissemination. Additionally, specific presentations used throughout the project will be made available under a CC-BY license version 4.0.

Regarding publications, the project's results will be disseminated electronically through publications in electronic channels, the press, or pertinent internet fora.  The project partners

will make an effort to communicate the project data that could be useful to external stakeholders while protecting the privacy of the stakeholders who will help with its development and collection. To accomplish the aforementioned goals, SmartCHANGE will securely deposit and keep its publicly accessible data on the Zenodo platform, which connects to OpenAIRE automatically to make the data accessible. Because of the accompanying metadata and Zenodo's search functionality, the data will be completely accessible. The majority of SmartCHANGE data released in Zenodo will take use of the Open Access option of data accessibility.

### 4.2.2   Openly accessible and closed data

The open access level will be used by SmartCHANGE to share project data with interested parties. Only authorized partners of the SmartCHANGE consortium will have access to the password-protected project repository where data that cannot be publicly disclosed will be kept. The template completed by each partner for each data they are going to collect/generate is available in Appendix I and Appendix II, whereas there is a specific section for them to outline the level of openness of these data ("Data access, sharing and licensing"). Public access to the open data will be ensured through inclusion of the data assets to the Zenodo platform and to the SmartCHANGE website.

### 4.2.3   Data anonymization

Only anonymized and aggregated data will be made public to ensure that data subjects cannot be identified in any documents (e.g., reports, publications), or data, within the project. The project partner, who is data controller in each instance, will carry out the necessary anonymization steps to ensure that the data cannot be used to identify the data subject. Consequently, during the process of anonymization, data identifiers need to be removed, generalized, aggregated, or distorted. The distinction between anonymization and pseudonymization must be made since the GDPR recognizes each as a separate category. Using the tour guide on data management from the Consortium of European Social Science Data Archives (CESSDA) as a source, Table 24 includes a list of best practices for the anonymization of quantitative and qualitative data.

*Table 24 – Data anonymization best practices*

| Type of data | *Best practices* |
|---|---|
| Quantitative data | • Reducing the accuracy or in-depth textual meaning of a variable, removing or aggregating variables<br>• Increasing or decreasing a variable's precision, such as age or residence. Report the least amount of georeferencing that will not likely jeopardize respondent privacy<br>• Generalizing the meaning of a descriptive text variable by substituting more generic text for potentially revealing free-text responses<br>• Limiting the upper or lower ranges of a continuous variable helps hide outliers if particular people's values fall beyond the norm for the larger group being studied |
| Qualitative data | • Editing identifying information using pseudonyms or general adjectives rather than leaving it blank<br>• Considering anonymization when transcribing or writing up the initial write-up (longitudinal research may be an exception if linkages between waves of interviews require special consideration for harmonised editing)<br>• Making use of uniform replacements or pseudonyms across the study team and the project. Using the same aliases in publications and subsequent research, for instance<br>• Carefully using "search and replace" procedures to avoid making unintentional changes and omitting misspelled words<br>• Clearly identifying substitutions in text, by using [brackets] or Extensible Markup Language (XML) elements<br>• Making an anonymization log of all replacements, aggregations, or removals made and storing it securely and apart from the anonymized data files (sometimes referred to as a de-anonymization key) |

### 4.2.4   Methods, software tools and documentation to access the data

There are no specific techniques, tools, software, or documentation that are intended to be used to access the data, regardless of the storage location or access policy.

## 4.3   Making data interoperable

By utilizing acceptable vocabularies (such as those for assigning search terms) along with proper standards for data and metadata production, it will be possible to make data *Interoperable*. The use of metadata vocabularies, standards, and methodologies that will improve the interoperability of the data collected/generated through SmartCHANGE's operations will be incorporated into its data management approach (in the data where it can be applied). Zenodo will support the interoperability of the data designated as open, with its

metadata being internally stored in a specific format (e.g., JavaScript Object Notation (JSON)) format in accordance with a predefined JSON schema.

The Dublin Core Metadata standard, which guarantees that the data meet conventional quality and consistency criteria while remaining interoperable with other data sources at the same time, will be studied in the case that it could be used by SmartCHANGE regarding the data that will not be shared with the general public - such action is under investigation and will be reported in M36. The standard comprises fifteen (15) metadata elements that present a vocabulary of concepts with definitions in plain language that may be transformed into processable open formats like XML, HTML, etc. Table 25 contains the Dublin Core Metadata standard's vocabulary.

*Table 25 – Vocabulary of the Dublin Core Metadata standard*

| No | Element | Element definition |
|----|---------|--------------------|
| 1 | Title | A name given to the resource. |
| 2 | Creator | An entity primarily responsible for making the content of the resource. |
| 3 | Subject | The topic of the content of the resource. |
| 4 | Description | An account of the content of the resource. |
| 5 | Publisher | An entity responsible for making the resource available. |
| 6 | Contributor | An entity responsible for making contributions to the content of the resource. |
| 7 | Date | A date associated with an event in the life cycle of the resource |
| 8 | Type | The nature or genre of the content of the resource. |
| 9 | Format | The physical or digital manifestation of the resource. |
| 10 | Identifier | An unambiguous reference to the resource within a given context. |
| 11 | Source | A reference to a resource from which the present resource is derived. |
| 12 | Language | A language of the intellectual content of the resource. |
| 13 | Relation | A reference to a related resource. |
| 14 | Coverage | The extent or scope of the content of the resource. |
| 15 | Rights | Information about rights held in and over the resource. |

## 4.4   Making data Reusable

A crucial part of the SmartCHANGE technique for making data FAIR is the reuse of data. SmartCHANGE will implement licensing to make sure that any interested stakeholder can make use of the publicly accessible data gathered/produced by the project. A license is a straightforward tool that allows a third party to copy, distribute, display, and/or alter the project's data as long as they do, so only for the license's predetermined uses. To encourage their reuse and provide an environment that is FAIR and accessible for them, SmartCHANGE will publish its freely accessible data under the CC licensing system (which is also supported

by Zenodo). The consortium believes that the CC BY-NC 4.0 licensing scheme is the most suitable one to ensure that the data is used as widely as possible while also ensuring that the participants' commercial interests are protected, and the source and authority of the data are acknowledged. Shortly, the latter restricts the use of the data so others cannot use it commercially or modify it in any way.

Prior to the publication of any data to Zenodo, SmartCHANGE will conduct Quality Assurance and Quality Control operations to ensure the transparency, consistency, comparability, completeness, and accuracy of the data. These initiatives will be carried out as input data checks, data summaries, and peer reviews.

The nature and substance of the data have a significant impact on its accessibility once the project is completed. Therefore, it is necessary to consult the data contributor before storing data on a public platform. The data tables in Section 3.6 offer information on the persistence of the data after the project's conclusion.

# 5 Allocation of resources

## 5.1 Costs for making data FAIR

The project budget includes the costs necessary to make the data gathered/generated during SmartCHANGE operations FAIR. These projected expenses will be required to pay for several specialized data processing and data management tasks, ranging from data collection and documentation through storage and preservation to sharing and reuse. The required effort will be included in the relevant Work Package (WP) as these activities are a part of the WP under which the relevant data are processed. The adoption of Zenodo will guarantee that the costs associated with long-term data preservation are reasonable.

## 5.2 Data management responsibilities

The responsibilities related to data management are specified as follows:

- **Work Package Leaders (WPL)** are in charge of organizing the data processing and quality control tasks carried out under the WP they are directing.
- **Task Leaders (TL)** operate as Data Controllers for the data gathered or created throughout the tasks they are in charge of. Additionally, they ensure that the data are correctly packaged for sharing with partners or for public consumption.
- The consortium partners who process the data created or collected are referred to as **Data Processors**. These processes involve data gathering, digitization, anonymization, storage, deletion, quality assurance, and consent acquisition.
- **Data Repositories** are intended for the long-term storage and preservation of project data. In earlier sections of this deliverable, Zenodo and the SmartCHANGE website were both discussed and described in relation to public data. As a result, the project's document repository that is only accessible by authorized project partners must safely store and maintain the data that can be shared among authorized consortium members during the course of the project.

# 6   Data security

Research ethics' main concern is data protection. All people have the right to control how information on them is gathered and used, which is protected by the EU Charter of Fundamental Rights.  Everyone has the right to the protection of their own personal data under Article 8(1) of the European Union's Charter of Fundamental Rights and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU), and the GDPR establishes guidelines for the protection of natural persons regarding the processing of personal data.

Anywhere that personal data - information pertaining to a named or distinguishable real person - is processed [EU 2016/679, 2023], extreme caution must be exercised. Data protection laws require researchers conducting study to give research participants comprehensive information about what will happen to the personal data they collect. Research involving sensitive data, such as health data, must receive special consideration because, per GDPR, such data cannot be processed without the subject's express consent or one of the other legal bases listed under Article 9(2) of the GDPR, such as processing "necessary for archiving purposes in the public interest scientific or historical research [...] [GDPR, Art. 9(2)(j)]".

According to the GDPR, the legal bases for processing personal data pertaining to stakeholders are that each data subject employed by a project partner has consented to the processing of his or her personal data (Article 6(1)(a)), and the processing is required for the performance of a contract, in this case, the employment contracts between the data subjects and each project partner (Article 6(1)(b)). Each project partner will keep records including the personal information of those who worked for them after the project is finished. Unless said employees prefer that their contact information be deleted, every partner is free to keep its copy of the contact information of employees who work for other SmartCHANGE partners. Only after the very last payment and evaluation from the EC, the mailing lists for the project will be removed. Only project participants will have access to the contact information of data subjects who are employees of any of the project partners, and only for as long as it takes to execute the GA and/or finish the project. Once the application is made publicly or commercially available, authorship information may be made public with the consent of the data subjects who are the authors.

All partners who process data have a responsibility to keep it secure by implementing all essential access controls (identification, authentication, and authorization) as well as security controls (including backup procedures and integrity checks) inside their infrastructure. In the

unfortunate event of a personal data breach, the project partners will promptly notify both the data subject(s) who may be impacted by the incident and their respective national supervisory authorities. At the same time, they will record any breaches of personal data and any relevant details.

If a data breach poses a risk to data subjects, a data controller must notify the supervisory authority within 72 hours of becoming aware of it [GDPR Art 4(12), 2023]. Any breach of personal data must be documented by the controller, describing the circumstances surrounding the breach, its effects, and the steps that were taken to address them.

In SmartCHANGE, regulatory compliance is viewed as a horizontal duty that affects all consortium members. All project participants are expected to make sure that the necessary steps are taken to prevent and lessen the effects of any incidents involving the security of the platform, the network, and the information systems. To ensure that the project's operations continue in a safe manner, all participants should execute an appropriate continuity strategy. To minimize negative impacts and stop similar incidents from happening in the future, it is important to evaluate the impact of each breach of personal information. When determining the significance of a personal data breach, a number of variables need to be taken into account, such as the quantity of individuals impacted, the severity of the impact on their rights and freedoms, as well as on project activities, and the length of the occurrence.

Each partner shall have a Data Protection Officer (DPO), where in case of a potential data breach, mitigation actions are considered in accordance with all the DPOs.

# 7 Ethical aspects

The implementation of SmartCHANGE will adhere to the highest ethical standards as well as applicable national, international, and EU laws governing ethical behaviour. While the ethical requirements are one of the reasons of WP2, whose goal is to ensure ethical compliance with the ethical standards stated in the project, ethics are significant in SmartCHANGE and are a distinct aspect of the GA.

This reflects to the report of the ethical standards of SmartCHANGE which corresponds to *D2.1 - Benchmark of regulatory and ethical frameworks*, which provides a description of key ethical and legal requirements, as well as their frameworks. This will guarantee that the project's technologies are in accordance with all applicable laws and regulations (e.g., GDPR), as well as any emerging requirements (e.g., AIA ).[1] Furthermore, D2.3, i.e. the *Societal, Legal, Ethical and Privacy (SELP) compliance framework*,  will deal with the contextual application to the SmartCHANGE project of the regulatory and ethical frameworks analysed in D2.1,, highlighting and  providing suggestions on how to reduce dangers and ethical issues related to the technology used in the project. The SmartCHANGE Ethical Board (EB) will be appointed within the context of deliverable D2.2 (i.e., *Ethics procedures for user engagement*) to check, on an annual basis, the research operations for compliance with ethical and legal requirements.

Another wave of questionnaires will be distributed to the consortium partners once the project's AI tools and software components have matured, with an emphasis on the EU AI Act (AIA) and the Assessment List for Trustworthy Artificial Intelligence (ALTAI) [ALTAI, 2023].

---

[1] See Veale et al. 2021 for a discussion on this proposed regulation

# 8 Conclusions

This version of the DMP of SmartCHANGE is the initial version of the DMP. It presents the project's data management policy in line with Horizon Europe open data requirements and defines the data management methodology of SmartCHANGE. A core aim of this methodology is to increase data FAIRness (i.e., to make data findable, accessible, interoperable, and re-usable to the widest possible extent).

Current technical approaches and standards are used by the SmartCHANGE DMP, including the OpenAIRE Zenodo research data repository, where many SmartCHANGE data will be merged. By using these tools, it will be possible to guarantee that the data created or gathered for SmartCHANGE will be saved, accessible, and usable long after the project is completed. However, some of the publicly accessible information and publications will also be made accessible via the SmartCHANGE website, which is frequently used by the active SmartCHANGE community.

The presented DMP aims to safeguard the confidentiality levels of the SmartCHANGE data, while ensuring adherence to GDPR mandates. Therefore, the DMP ensures the application of the necessary informed consent processes.

This version of the DMP is delivered in the 6th month of the project, and as a result it is considered a living document and will be updated in the next months, by the time that new data become available, or a need for changes in the data management methodology arises.

# 9    References

AIA: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Assessment List for Trustworthy Artificial Intelligence (ALTAI) (2023), https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

DataCite (2023), https://datacite.org/

European Commission - Turning FAIR Into Reality (2023), https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

Florio P, Giovando C, Goch K, Pesaresi M, Politis P, Martinez A. TOWARDS A PAN-EU BUILDING FOOTPRINT MAP BASED ON THE HIERARCHICAL CONFLATION OF OPEN DATASETS: THE DIGITAL BUILDING STOCK MODEL-DBSM. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2023 Jun 22; 48:47-52.

GDPR: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (general Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Gobbato S. Open Science and the reuse of publicly funded research data in the new Directive (EU) 2019/1024. Journal of Ethics and Legal Technologies. 2020 Nov;2(JELT-Volume 2 Issue 2):145-61.

Go FAIR - FAIR Principles (2023), https://www.go-fair.org/fair-principles/

Horizon Europe - Data Management Plan Template (HE DMP) (2023), https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf

Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, Courtot M, Crosas M, Dumontier M, Evelo CT, Goble C. FAIR principles: interpretations and implementation considerations. Data intelligence. 2020 Jan 1;2(1-2):10-29.

Johnson G. Economic research on privacy regulation: Lessons from the GDPR and beyond. 2022.

Ulrich H, Kock-Schoppenhauer AK, Deppenwiese N, Gött R, Kern J, Lablans M, Majeed RW, Stöhr MR, Stausberg J, Varghese J, Dugas M. Understanding the nature of metadata: systematic review. Journal of medical Internet research. 2022 Jan 11;24(1):e25440.

Veale M, Zuiderveen Borgesius F. Demystifying the Draft EU Artificial Intelligence Act - Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International. 2021 Aug 1;22(4):97-112.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016 Mar 15;3(1):1-9.

Zenodo (2023), https://zenodo.org/

# Appendix I.   Data information template

Data Information Template (completed by each partner).

| Data name | Description |
|---|---|
| Data description | |
| Related WP/Task | |
| Data origin | |
| Will you re-use any existing data? If yes, how? | |
| Methodologies for data collection/generation | |
| Data format | |
| Where this data will be stored? | |
| Expected size of the data | |
| Metadata and standards used | |
| For whom might the data be useful? | |
| Data access, sharing and licensing | |

**Funded by the European Union**

# Appendix II.  Pilot data information template

Data Information Template (completed by each pilot partner)

| *Data name* | *Description* |
| --- | --- |
| Data description | |
| Related WP/Task | |
| Data origin | |
| Will you re-use any existing data? If yes, how? | |
| Methodologies for data collection/generation | |
| Data format | |
| Where this data will be stored? | |
| Expected size of the data | |
| Metadata and standards used | |
| For whom might the data be useful? | |
| Data access, sharing and licensing | |
| **Additional information** | |
| Is the data personal? | |
| Is the data sensitive? | |
| Is there any Data Protection Officer (DPO) responsible for this data? If yes, provide contacts | |
| Is there any data protection policy assigned to this data? If yes, provide a copy | |