Addressing Psychosocial and Lifestyle Risk Factors to Promote Primary Cancer Prevention: an integrated platform to promote behavioural change (iBeCHANGE)

Project Number: 101136840

D7.1 – iBeChange Data Management Plan

Related Work Package	**WP7 – Ethical, Privacy, and Data Protection**
Related Task	**Task 7.1 – Ethical and Legal Inventory**
Lead Beneficiary	i~HD
Contributing Beneficiaries	IEO, ICO, EUT, SPD, UNIPA, POLIMI
Document version	V1.0
Deliverable type	DMP
Dissemination level	PU
Due date	31/05/2024
Delivery date	29/05/2024
Authors	Nathan Lea (i~HD), Jens Declerck (i~HD), Dipak Kalra (i~HD), Muna Khogali (i~HD)
Contributors	Aline Machiavelli (SD), Carolina Migliorelli & David Sunol (EUT), Francesco Trovò & Emilia Ambrosini (POLIMI), Dario Monzani (UNIPA), Anna Garcia Serra (ICO), Giorgia Miale (IEO), Massimo Monturano (IEO), Gabriella Pravettoni (IEO), Marianna Masiero (IEO)
Reviewers	Reviewed by all consortium

Funded by the European Union

This project has received funding from the European Union’s Horizon Europe research and innovation programme under the Grant Agreement Number 101136840.

Disclaimer

Version history

Version	Description	Date completed
V0.5	Initial draft for Consortium Input form Nathan Lea	24/04/2024
V1.0	Submission Draft for Management Board Review after partner checks, cocreation meeting and questionnaire initial updates.	10/05/2024

Table of Contents

List of Abbreviations

Abbreviation	Explanation
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Reusable Principles
RCT	Randomised Control Trial
GDPR	General Data Protection Regulation
DPIA	Data Protection Impact Assessment
REC	Research Ethics Committee
EHDS	European Health Data Space
MDR	Medical Device Regulation
HTA	Health Technology Assessment
PROMS	Patient Reported Outcome Measures

List of Tables

List of Figures

Executive Summary

Deliverable 7.1 provides the first Data Management Plan (DMP) for iBeChange. A Data Management Plan is an important tool for helping multi-party consortium projects like iBeChange to develop a consensus view early in the project of how data will flow across the partners and how it will be handled to achieve the goals of the project.

This forms part of the overall information governance approach and framework to ensure that iBeChange remains compliant with the General Data Protection Regulation (GDPR) and other regulations, including research governance and forthcoming regulations including the AI Act.

The DMP is helping to develop the Data Protection Impact Assessment (DPIA) for iBeChange as part of that compliance. Both DMP and DPIA have been informed by not only an initial assessment of the project’s goals and proposed activities as specified in the Grant Agreement, but also a co-creation workshop around the development of the platform held on 26 April 2024.

A series of questionnaires sent to partners focusing on data handling in April and May 2024 have also helped to inform the DMP. Responses are still being collected and the questionnaires have been designed to support the DPIA. The questionnaire template is available under Annex 1 in this deliverable and will also be made available along with responses in D7.2 scheduled for M12.

A DMP further provides an opportunity for a consortium to take a view on the security of the data and how it will be stored, and also whether it will be made available for wider use in line with the Open Science Agenda and the Findable, Accessible, Interoperable and Reusable (FAIR) Principles.

In the case of iBeChange this first DMP is published as a basis upon which the other risk management and information governance processes can proceed. The DPIA and further assessments will inform the development of data sharing agreements and development of codes of practice for the project by M18.

1. Introduction

This Deliverable is based on the Data Management Plan template provided by the European Commission for Horizon 2020 projects[1]. A Data Management Plan is an important tool for helping multi-party consortium projects like iBeChange to develop a consensus view early in the project of how data will flow across the partners and how it will be handled to achieve the goals of the project.

The DMP forms part of the overall information governance approach and frameworks to ensure that iBeChange remains compliant with the General Data Protection Regulation (GDPR) and other regulations, including research governance and forthcoming regulations including the AI Act. In addition, iBeChange partners must also follow their own local procedures for regulatory compliance. These include a need to demonstrate ethical adherence (including Research Ethics Committee (REC) approvals at participating sites) and ensure that the appropriate consents and information leaflets are made available to participants.

iBeChange must also consider a wider set of existing and forthcoming regulations that will impact the project. The current regulations may include the Medical Device Regulation that imposes a series of Health Technology Assessment (HTA) needs for any interventional tool. The forthcoming regulations including the AI Act and the European Health Data Space (EHDS). The AI Act will bring about a series of certifications for high-risk systems, including those with medical decision making and access to services intervention. Along with the EHDS, the AI Act imposes requirements around data quality and documentation for the certification process.

Adherence to these regulations is forming part of the overall strategy for compliance being led by WP7. The DMP provides a foundation upon which to document initial plans for handling data and a basis to proceed with the compliance. The DMP is helping to develop the Data Protection Impact Assessment (DPIA) for iBeChange as part of that compliance. Both DMP and DPIA have been informed by not only an initial assessment of the project's goals and proposed activities as specified in the Grant Agreement, but also a co-creation workshop around the development of the platform held on 26 April 2024.

A series of questionnaires sent to partners focusing on data handling in April / May 2024 have also helped to inform the DMP. Responses are still being collected and the questionnaires have been designed to support the DPIA. The questionnaire template is available under Annex 1 in this deliverable and will also be made available along with responses in D7.2 scheduled for M12.

A DMP further provides an opportunity for a consortium to take a view on the security of the data and how it will be stored, and also whether it will be made available for wider use in line with the Open Science Agenda and the Findable, Accessible, Interoperable and Reusable (FAIR) Principles. iBeChange is not planning to make data available for further

[1] Full template available here for reference: https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx

reuse outside the project at this stage. As discussed later in the DMP, the project will review this position.

In the case of iBeChange, this first DMP is published as a basis upon which the other risk management and information governance processes can proceed. The DPIA and further assessments will inform the development of data sharing agreements and development of codes of practice for the project by M18.

An overview of the project and its pilot study is provided in figure 1 below.

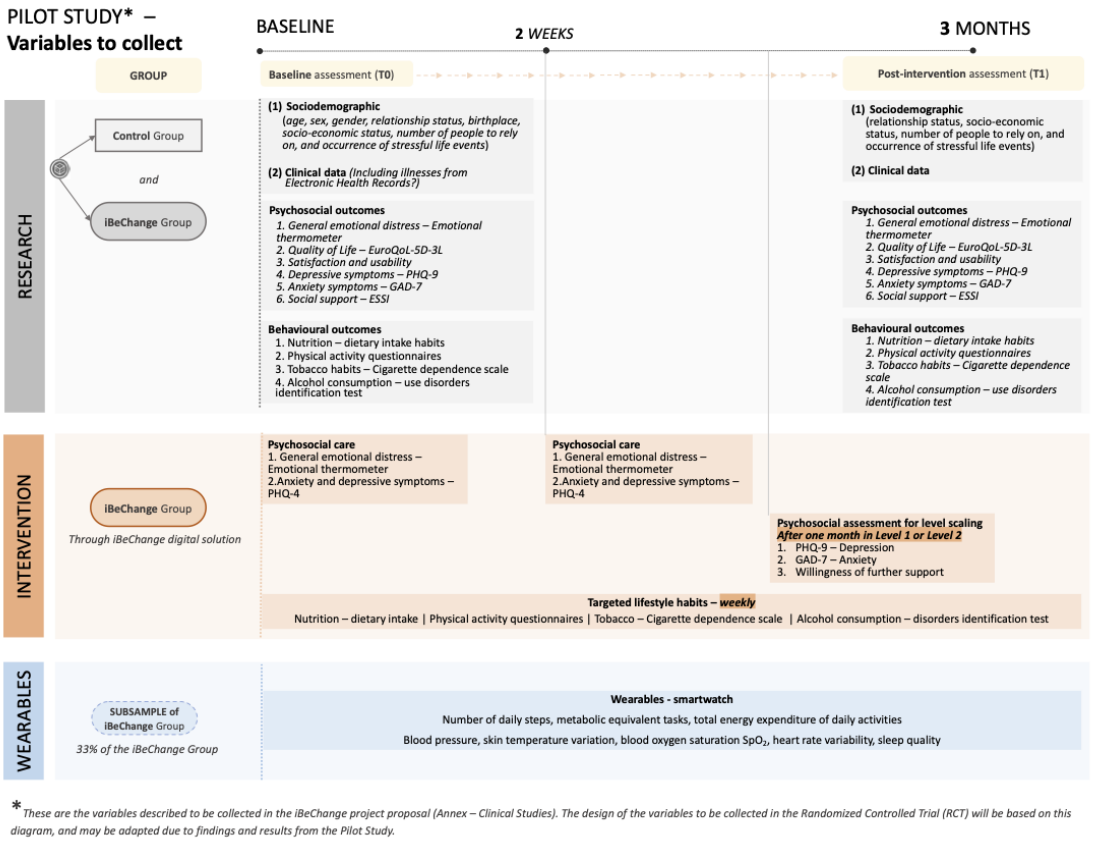

Figure 1: ICO Pilot Study Variables Presentation

2. Data Summary

Aspect	Response/explanation
Purpose of the data collection/ generation and its relation to the objectives of the project	iBeChange is developing an integrated platform that is attempting to promote healthy lifestyles and the requisite behaviour changes to minimise the risk of developing cancer in the population. The Project will be reviewing existing cancer histories to develop a basis for recommendations to be made on behaviour changes based on the experiences of those who have developed cancer. The data will also involve questionnaires for participants in terms of their behaviours and their lifestyle factors to help develop the behaviour change heuristics in line with behaviour change theory. It will include psychosocial and sociodemographic data form the outset, and throughout the research projects and interventional trials a follow up collection of psychosocial and sociodemographic factors will be assessed. Wearables and environmental data will also be used to inform the platform processes for a subset of the participating cohort. The data collected is therefore related to cancer care and will be used to train algorithms that will help tailor recommendations to users on how to manage their activity and lifestyle. The intervention will be trialled in a pilot study and Randomised Control Trial

Types and formats of data generated or collected by the project	• Behavioural profile data (including sociodemographic and psychosocial data)
	  ○ Eye movements, gaze behaviour, and pupil dilation for the eye tracking study
	• Location and wider environment data
	• Wearable device data
	• Demographics and sociodemographic
	  ○ Including age, gender, education, marital status, household, employment, and income
	• User registration and management
	• Clinical and outcomes data
	• Risk management and modelling data for unhealthy risk factors
	• Emotional and wellbeing data
	• User feedback and initial questionnaires

	○ *Qualitative (e.g., focus groups and semi-structured interviews) and quantitative (e.g., through self-report measures) information about acceptability, acceptance, usability and user experience* • Self monitoring, clinical care support • eCRF trials data • Engagement and activity data including via wearables and sensors ○ heart rate, skin conductance, blood pressure, oxygen saturation, steps taken, sleep patterns, and calories burned ○ Objective measures of user engagement (e.g., logs) • Patient Reported Outcome Measures (PROMS) • Voice data for the analysis of stress and emotions • Participant interaction with the application data (clicks, pages viewed etc.) • Wearables data will be transferred via JSON.
Any re-use existing data and how this will be done	Existing clinical data for participants will be accessed to be able to train the behavioural and outcomes heuristics (AI models) for the platform in assessing risks of unhealthy lifestyles, interacting with the participants through use and gamification.
The origin of such data	Clinical records form care providers in participating centres.
Expected size of the data	Uncertain at this point – likely no more than 1 GB per participant (assuming no image data is to be used).
Likely users of the data	• Platform developers and AI Model trainers; • Behavioural change researchers; • Heuristics programmers • Research Participants • Trial Managers at the Sites • (Sustainability - Competent Authorities for Certifications around MDR and AI Act compliance)

Table 1: Data Summary

3. FAIR Data - making data findable, including provisions for metadata

Aspect	Response/explanation
Are the data produced and/or used in the project discoverable, identifiable and locatable by means of a standard identification mechanism	• For the existing clinical data: this depends on the data management in each of the participating sites. • For the data collected for the project including in questionnaires and user feedback: the WPs responsible for developing these will need to catalogue the data in accordance with requirements for conducting interventional trials. This needs to be further specified moving forward as the platform and interventional aspects develop.
What standard identification mechanism used (e.g. persistent and unique identifiers such as Digital Object Identifiers)	None identified as of yet but DOIs should be defined.
Is meta-data available through catalogue?	There are no plans to develop a catalogue of data but cataloguing individual items is recommended as part of Data Quality Assessment Readiness Planning and will be explored.
Can meta-data be used for search?	This will need to be decided as part of the project development – likely a common standard would be used should a catalogue be developed.
Naming conventions used	TBC – not currently applicable
Clear versioning supported?	This will need to be implemented to assist with AI Act and MDR certifications but also for the development process over the next 18 months
Additional keyword search supported?	TBC – not currently applicable
What metadata will be created using which standards?	TBD - OMOP? LOINC? This is a diverse data set so may need to look beyond medical terminologies and metadata should it be made available outside of the project.

Table 2: FAIR data- Making data findable, including provisions for metadata

4. Making Data Openly Accessible

Aspect	Response/explanation
Will data be made openly available as the default?	No. The data collected will be useful for only the development of the platform and its validation through the trials and participant interactions. The Project will consider open availability based on utility and appropriate consents as the programme develops
Which datasets will NOT be openly available and why?	Currently none - but see above for utility and regulatory needs,
How will the data & meta-data be made accessible (e.g. by deposition in stated repository)?	TBD if applicable.
If known repository, what arrangements explored?	Not yet known
If project-specific access, then:	
Data Access Committee	The work will be overseen by the appointed Ethical Advisory Board of experts and in line with REC approvals. Should data be reused an appropriate Data Access Committee will be established as part of a sustainability approach.
Any conditions for access (i.e. a machine-readable licence)	There are no plans to licence data reuse at this point. A creative commons licence or commercial fee-based licensing may be considered.
What methods or software tools will be needed to access the data?	There are no plans to make the data downloadable however solutions will focus on the possibility of using GITHub secured or other repository
Documentation for software	Software documentation will be rigorously adhered to for future AI Act and MDR compliance as well as the development paradigms for interface and machine learning. This will need to be in a programme management tool (e.g. JIRA or Confluence) along with the code that it relates to. This will meet the standard for ay software reuse moving forward and sustainability
Availability of software	See above - no plans other than for appropriate licensing after trials completion

Institution and researcher vetting process/procedures - describe	All processes will follow the 5 safes for the project, including safe people. This entails training and contractual obligations on the employee so that they can work within the partner institutions. We will also look to develop a Code of Practice in line with GDPR for researchers and institutions (reference D7.4). All institutions need to hold appropriate insurance for participating in the project as well as entering into Data Sharing Agreements (being defined using the DPIA process that is currently underway). Partners can also vet individual access needs on a per case basis.

Table 3: Making Data Openly Accessible

5. Making Data Interoperable

Aspect	Response/explanation
Are the data produced in the project interoperable	The data will need to be interoperable within the platform itself but any data derived for future use may be shareable using interoperability standards (if appropriate). For sustainability purposes the data will need to be sharable for certification.
If not, explain why not	See above.
Data and metadata vocabularies, standards or methodologies used	Not yet decided - may not be appropriate
Standard vocabularies used	Possibly from the clinical source data only dependent on each site’s standards
Mappings from uncommon or project-specific ontologies or vocabularies to more commonly used ontologies	None foreseen. Please note that behavioural change data items may not be labelled or categorised as standard outside of e.g. SNOMED CT.

Table 4: Making Data Interoperable

6. Increase Data Re-use (through clarifying licences)

Aspect	Response/explanation
Will data be available for onward data-sharing/re-use?	Unlikely. Data is useful for development of the platform and trial purposes only for this project. This will be assessed throughout the project lifetime
Approach to data licensing for onward use	TBD. Licencing will be considered in line with sustainability plans should the data itself demonstrate reuse utility outside of the project and platform operation.
Likely date for data availability for onward use	TBD – not currently applicable.
Explain any restriction on date of availability	N/A
Possible restrictions on onward data-sharing	Confidentiality and privacy dependent on outcomes of ethical reviews.
Data retention policy (including availability for data-sharing)	In line with regulatory retention for trial compliance and research governance. This will also include any: • AI Act certification needs; • Medical Device Regulation; • Health Technology Assessments.
Description of data quality assurance processes	Please refer to i~HD Data Quality Checklist Plan in the DPIA. This will be published as part of D7.2.

Table 5: Increase Data Reuse (through clarifying licenses)

7. Allocation of Resources

Aspect	Response/explanation
Estimated project costs for making data FAIR	TBD if applicable
Data management responsibility across the project	Currently shared between IEO (Istituto Europeo di Oncologia) and i~HD working with each partner representative
Resources required for long term preservation (costs and potential value, who decides and how what data will be kept and for how long)	TBD if applicable

Table 6: Allocation of Resources

8. Data Security

Aspect	Response/explanation
Data security measures used (including data recovery as well as secure storage and transfer of sensitive data)	TBD as part of the DPIA process and Platform Development work. It will focus on: • Central Platform Security and data access; • User Access credentials for the Interfaces (likely MFA)
Where data will safely be stored (in certified repositories for long-term preservation and curation). Provide detail	At current publication this has not yet be determined – institutes are looking at their existing setups including Amazon Web Services and this will be further determined and overseen by the DPIA.

Table 7: Data Security

9. Ethical Aspects

Aspect	Response/explanation
Any ethical or legal issues that can have an impact on data sharing	Legal issues along with ethical compliance are being explored and assessed with the DPIA. This will continue throughout the project. Independent Research Ethics Committees will review the proposals and take a view on any specific challenges The iBeChange EAB will oversee and identify specific challenges as they meet and review. Expected challenges involve: • Accuracy and reliability of risk identification and recommendations from the system (AI related) • Risk of distress caused by behavioural change attempts • Digital divide challenges such as requirements for tech savvy users and participants
References to ethics deliverables and ethics chapter in the Description of the Action (DoA) – if relevant	Reference the following: • WP7 for ELSI • Deliverables D7.3, D7.6 and D7.8 for the EAB reports. • The Information Governance Framework D7.4
Questionnaires dealing with personal data	As above - risk of distress caused in considering sensitive and upsetting topics, as well as behavioural change
How is informed consent for data sharing and long-term preservation sought in such questionnaires?	This will form part of a trials protocol that is under development for the interventional studies, ie: • Informed Consent will be achieved via recruitment in sites • detailed information leaflet as part of a Study Information Package. • All materials will be reviewed by the RECs at each participating site and amendments will be made where necessary.

Table 8: Ethical Aspects

10. Sustainability and Future Compliance Risk Management

Aspect	Response/explanation
What elements are planned for sustainability activities moving forward?	The sustainability plans are being developed and will evolve as the project proceeds. Reference the following: • WP8 • Deliverables D8.6 Sustainable policy strategy
What certifications would be required for a sustainability plan to go to market? (e.g. MDR, AI Certifications)	Certifications may include, and not be limited to, the following: • MDR 2023 (Medical Device Regulation) • HTA (Health Technology Assessments) • AI Act 2024
What data retention and best practice documentation is required to achieve this? Specifically, how can we ensure the quality, assessment of bias, traceability of results and recommendations, transparency, repeatability etc.	Refer to ALTAI (Assessment List for Trustworthy AI) requirements: • human agency and oversight • technical robustness and safety • privacy and data governance • transparency • diversity, non-discrimination and fairness • environmental and societal well-being and • accountability
What risk management is envisaged to support sustainability? E.g. DPIA Updates, Ethical Review, Tool management, Legacy System reliance etc.	The DPIA will be continually updated to support any sustainability plans. Where appropriate RECs will be asked to consider sustainability plans as part of independent oversight and participant consent. The EAB will also provide views on sustainability.
Is there a data flow diagram available?	See figure 1 for an initial high level diagram but this is to be developed with the DPIA process.

Table 9: Sustainability and Future Compliance Risk Management

11. Conclusions

Deliverable D7.1 provides the first version of the DMP for iBeChange. The DMP will be updated later in the project. D7.1 offers a snapshot view of the planning around data management as at M6 of the project. These details are subject to change and update, and these will be reflected in the updated plan due in M36 of the project as part of D7.7.

The next steps for the DMP involve its use in assisting the DPIA process. The DPIA itself will go into further detail around the particulars of data processing across the partners. As the development process continues and any amendments are needed to the planning as described here, the amendments will be risk managed as part of the DPIA process and updated in M36.

12. Appendix – Initial Data Handling Questionnaire Template

Information Governance Questionnaire for iBeChange

This questionnaire will form the basis for the Data Protection Impact Assessment (DPIA) and Data Flow diagrams.

Please provide as much detail as you can at this time. If you have any questions or concerns, or require further clarifications, please contact nathan.lea@i-hd.eu.

We anticipate that this questionnaire will take no more than 30 minutes to complete and would be grateful if you can complete this questionnaire by close of business on Friday 3rd May.

If you have not yet finalised the details for your data processing please say so in your responses and give an anticipated date for completion.

i~HD is collecting the responses and will process the data in accordance with their transparency notice available at https://www.i-hd.eu/data-protection-transparency-and-privacy-notice/.

Name of institute you are working for:

Your preferred contact email address:

Who is completing the questionnaire?

Aspect	**Response/explanation**
Are you a Data Provider, Recipient, both or something else?	[Confirm whether you are a Data Provider partner or a Recipient, and what any other role you have ie both Data Provider and Recipient, or any other role]
Which Work Packages (WPs) and parts of the project are you working on?	[Provide a list of the WPs that you are working on. Delete and add to as necessary brief detail of what you are working on ie Trials, Development, algorithm training etc.] • WP1 - Project management and coordination • WP2 - Design of the integrated platform using Behaviour Change Techniques and User-Centered Design approach • WP3 - Novel approaches for interaction through a Virtual User Model • WP4 - Intelligent Reinforcement Learning-based Personalised Behaviour Change Platform • WP5 - Deployment of the Recommendation System to the population • WP6 - Impact assessment and policy-making

	• WP7 - Ethical, Privacy, and Data Protection • WP8 - Policy Implementation, stakeholder/policy dissemination, exploitation and communication • WP9 - Ethics Requirements
What existing data do you intend to share, need to receive and/ or need to capture? (N.B. see next question about for new data collection)	[Provide a list of data. Below is an example of a list to delete and add to as necessary.] • Clinical data • Images • Wearable data • Participant Questionnaire Responses • Socio, psychosocial data • Other please specify
What data will be collected from participants?	[Provide a list of data types. Below is an example of a list to delete and add to as necessary.] • Behavioural data (please specify) • Sociodemographic data (please specify) • Location and wider environment data • Wearable device data • Demographics • User registration and management • Psychosocial data (please specify) • Clinical and outcomes data • Risk management and modelling data for unhealthy risk factors • Emotional and wellbeing data • User feedback and initial questionnaires • Self monitoring, clinical care support • eCRF trials data • Engagement and activity data
What formats are you expecting to share with / receive using?(This question will depend on what the researchers and architects need and the decisions on any eCRFs and activities within the platform as well	[Provide a list of data formats. Below is an example of a list to delete and add to as necessary.] • EHR Standard (e.g. HL7, FIHR) • CSV

as what data sites are able to share).	• OMOP • DICOM • STATA / • Other please specify
Who do you intend to share the data / expect to receive data from?	[Provide a list of potential users and their organisations. Below is an example of a list to delete and add to as necessary and any named users if relevant.] • Platform developers and AI Model trainers; • Behavioural change researchers; • Heuristics programmers • Research Participants • Trial Managers at the Sites • (Sustainability - Competent Authorities for Certifications around MDR and AI Act compliance
What will recipients do with data when they receive it and how will they process it? Will there be any onward sharing with other partners?	[Provide detailed description and upload any relevant images or documents]
What system architecture will be used for managing the Pilot and the RCT?	[Provide detailed description of the hardware and software components and provide any available drawings]
Do data providers intend to anonymise or pseudonytmise the data? (This may be needed under GDPR compliance)	[Provide detailed description]
Do you intend to make any data available for wider use outside of iBeChange? (We need answers to these questions by Mid May)	[Yes (Describe and specify what, how and why) No or N/A (Describe if relevant) TBD (include time frame of decision if known)]

Please add below any other information you think is relevant.