This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101007990.

Project Acronym:	**VITALISE**
Project Title:	Virtual Health and Wellbeing Living Lab Infrastructure
Project Number:	101007990
Topic:	**Horizon 2020 Research and Innovation Programme** **INFRAIA-02-2020** **Integrating Activities for Starting Communities**
Type of Action:	**RIA - Research and Innovation action**
Start date of the Project:	**April 2021**
Duration of the Project:	**36 months**

D1.4 Data Management Plan (first version)

(Version 3.0, 29/09/2021)

Deliverable:	D1.4 Data Management Plan (first version)	
Work Package:	WP1 Management	
Due Date:	30/09/2021	
Submission Date:	30/09/2021	
Lead Beneficiary:	UPM/AUTH	
Version:	3.0	
Status:	Submitted	
Author name(s):	Panagiotis Kartsidis (AUTH), Alexandra Anagnostopoulou (AUTH), Leidy Enriquez (ENoLL), Alejandro Medrano Gil (UPM)	
Contributor(s):	Despoina Petsani (AUTH), Sylvie Bernaerts (LICALAB), Teemu Santonen (LAUREA)	
Reviewer(s):	Idoia Muñoz (GAIA)	Gorka Epelde (VICOM)
Approved by:	Evdokimos Konstantinidis, Project Coordinator	
Keywords:	Data management, handbook	
Nature:	☒ R – Report ☐ P – Prototype ☐ D – Demonstrator ☐ O - Other	
Dissemination level:	☒ PU - Public ☐ CO - Confidential, only for members of the consortium (including the Commission) ☐ RE - Restricted to a group specified by the consortium (including the Commission Services)	

The **VITALISE** Consortium consists of:

Participant #	Participant organisation name	Short Name	Country
1 (Coordinator)	EUROPEAN NETWORK OF LIVING LABS IVZW	ENoLL	Belgium
2	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTH	Greece
3	LAUREA-AMMATTIKORKEAKOULU OY	LAUREA	Finland
4	THOMAS MORE KEMPEN VZW	LICALAB	Belgium
5	FUNDACION INTRAS	INTRAS	Spain
6	TREBAG SZELLEMI TULAJDON- ES PROJEKTMENEDZSER KORLATOLT FELELOSSEGU TARSASAG	TREBAG	Hungary
7	ASOCIACION DE INDUSTRIAS DE CONOCIMIENTO Y TECNOLOGIA - GAIA - EUSKALHERRIKO EZAGUTZA ETA TEKNOLOGIA INDUSTRIEN ELKARTEA	GAIA	Spain
8	FUNDACION CENTRO DE TECNOLOGIAS DE INTERACCION VISUAL Y COMUNICACIONES VICOMTECH	VICOM	Spain
9	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	CERTH	Greece
10	AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH	AIT	Austria
11	SOCIALIT SOFTWARE E CONSULTING SRL	SIT	Italy
12	UNIVERSIDAD POLITECNICA DE MADRID	UPM	Spain
13	WITA SRL	WITA	Italy
14	ANTHOLOGY VENTURES JSC	AV	Bulgaria
15	VILABS OE	VILABS	Greece
16	FORUM DES LIVING LABS EN SANTE ET AUTONOMIE	LLSA	France
17	TECHNISCHE UNIVERSITEIT INDHOVEN	TUE	Netherlands
18	ROYAL INSTITUTION FOR THE ADVANCEMENT OF LEARNING MCGILL UNIVERSITY	McGill	Canada
19	UNIVERSITE DE MONTREAL	UDEM	Canada

Revision history			
Version	**Date**	**Modified by**	**Comments**
V0.0	01/07/2021	ENoLL, AUTH	Table of contents
V0.1	31/08/2021	AUTH	Introduction, ethical aspects
V0.2	05/09/2021	AUTH	Data security, Data anonymization
V0.3	15/09/2021	AUTH	Data surveys, Data summary
V1.0	23/09/2021	AUTH	Document upgraded for review
V1.2	27/09/2021	AUTH	Document updated following internal review suggestions
V2.0	28/09/2021	EnoLL	Document upgraded to final review version
V2.1	29/09/2021	AUTH	Adressed final review comments
V3.0	30/09/2021	EnoLL	Document upgraded to final version for submission

TABLE OF CONTENTS

List of Tables

s

Abbreviations

Acronym	Description
API	Application Programming Interface
DMP	Data Management Plan
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
H2020	Horizon 2020 Program of the European Commission
ID	Identity document
JRA	Joint Research Action
N/A	Not Applicable
OpenAIRE	Open Access Infrastructure for Research in Europe
PGH	Participant Generated Health Data
TBD	To be Defined
QoL	Quality of Life
WP	Work Package

Executive summary

In line with the *Open Research Data Pilot* regulations of the Horizon 2020 programme (H2020) (Grant Agreement Number 101007990 - VITALISE, 2021), the purpose of this deliverable is to present the first version Data Management Plan (DMP) of the VITALISE project (Open Access - H2020 Online Manual, 2017). The document provides provisory information regarding the data to be collected during the project, as well as addresses the issue of data re-use for further exploitation by expanding on the manner in which the data will be made accessible and gives information on the data's curation and preservation.

According to the FAIR (Findable, Accessible, Interoperable and Reusable) Data Management guidelines in Horizon 2020 and in agreement with the General Data Protection Regulation (EU) 2016/679 (GDPR) (General Data Protection Regulation (EU GDPR) 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, 2016), all datasets to be produced and collected and shared amongst the partners during the project were identified and an first version of a list containing all relevant details was established.

Currently the project is set to develop and share a collection of datasets related to patient rehabilitation supported by technology, transitional care and everyday living environments. These datasets will also be uploaded to the VITALISE platform that will facilitate open access data processing allowing for federated and harmonized living lab data access.

Considering the fact that the VITALISE datasets will at their majority be produced and collected from human participants, the respective data (raw or processed) should be managed in total agreement to the guidelines and regulations in place regarding ethical and privacy issues related to datasets of this nature.

This document is the first version of the VITALISE DMP submitted at month M06 of the project. Consequently, the datasets presented at this stage represent a preview of the data that may be collected. In the later stage of the project, we expect that there may be some changes regarding the content of the datasets. Important changes occurring due to the inclusion or exclusion of a dataset will be followed by changes in consortium policies or external factors and the creation of an updated version of the DMP. To ensure that the DMP will be implemented as expected regular checks on the status of the data will be carried out on a schedule. The main principals of the project, though, are expected to be maintained as is until the end, thus delineating the principal axes of the overall DMP.

The final version of the DMP will be submitted at month M30 of the VITALISE project in deliverable D1.5 "Data Management Plan (final version)".

1. Introduction

1.1. Purpose of the document

This document describes the first version of the Data Management Plan (DMP) of the EU funded project VITALISE (Horizon 2020 Programme, under Grant Agreement N. 101007990). The purpose of this document is to:

1. present information regarding the datasets to be generated, collected, and processed in the duration of the VITALISE project;
2. specify the data management policy to be applied in relation to these datasets;
3. describe the methods and standards that will be implemented for these datasets;
4. specify the manner in which the data will be shared, made open, curated, and preserved;
5. provide an overview of the measures deemed necessary for satisfactory data management from an ethical and security perspective.

The VITALISE DMP stipulates the measures for advancing the findings of the project and establishes the processes for sharing research and non-research data. To that aim, this document follows the structure of the Horizon 2020 DMP template (H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, 2016). In accordance with the template, it reports on the data that will be generated, collected and processed, and the manner in which the data will be shared, made open, curated and preserved, guaranteeing open access by following the adequate licensing scheme (e.g. Creative Commons License).

Definitions of the general policy and the approach to data management in VITALISE will also be given in this first version of the DMP. This encompasses data and meta-data collection, publication and deposition of open data, the infrastructure of data repository and compliance to the Open Access Infrastructure for Research in Europe (OpenAIRE) (OpenAIRE, 2017).

Finally, the DMP details the intermediate results from activities concerning data collection, which are fulfilled following the FAIR (Findable, Accessible, Interoperable, and Reusable) data management (H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, 2016). The DMP will be updated within the duration of the project. The updated versions will refine aspects of the policy and provide further details about the generated and collected datasets of the VITALISE project.

1.2. Grant Agreement provisions

In the VITALISE Grant Agreement we have made provisions regarding the contents of the DMP, as well as open access policy. Specifically, in pages 57 and 58 of ANNEX I (PART B)(Grant Agreement Number 101007990 - VITALISE, 2021) the following are stated:

*"**Data Management:** VITALISE is a proponent of the Open Data Initiative and will ensure that the collected data will be made available to the public (WP3), as part of the provided access to research infrastructures, through its Virtual Access to datasets mechanism (WP12) provided that privacy and security issues are addressed, adhering to legal and ethical guidance for handling personally identifying data and be in line to the GDPR regulation. In order to do that, we will work on a Data Management Plan (DMP) quite early in the project (i.e., D1.2, and D1.5- M6), which will determine the type of generated data, as well as their exploitation and curation strategy that will be monitored and revised based on project's needs. More specifically, a general outline of the DMP consists in: a) Specifying the data that will be collected through the JRAs monitoring and intervention; b) Investigating best practices and guidelines for working with Open Data available by organisations or institutions supporting Open Data initiatives. These standards will be taken into account and include the ones released by the: i) Open Data Foundation, and ii) Open Knowledge Foundation; and c) Defining how the data collected in the project will be made available to third parties in contexts, such as scientific scrutiny, peer review and use for research or commercial purposes. To this end the VITALISE platform architecture goes beyond the current state of the art and facilitates data access.*

To maintain adherence to guidelines for the management and protection of personally identifying data, a nominated data controller role within the Project Steering Board (PSB) has also been included. This will include ensuring any personal data that are collected or stored by the project adheres to the principles of informed consent for a clear and specific purpose, using secure storage, with right of access; and where data collected is adequate; it is relevant and not excessive. DMP will identify best practises and specific standards for these major data types and assess their suitability for sharing and reuse in accordance with official guidelines. Identified datasets for sharing will be made discoverable, accessible, assessable and intelligible, useable and interoperable to specific quality standards, using data sharing services such as OpenData (open- data.europa.eu/en/data/) or EUDAT (www.eudat.eu/) for open access."

With regards to the open access policy, it is stated in page 57 of the Grant Agreement ANNEX I (PART B) (Grant Agreement Number 101007990 - VITALISE, 2021) that:

"Moreover, in order to maximise project knowledge, participants will prepare publications based on their work in the project. Steps will be taken to ensure open access to peer-reviewed publications in line with the requirements of Horizon 2020. The consortium will aim to work in line with the principles of the ***Open Research Data Pilot****. The standard approach will be to follow the 'green' route to open access, i.e. copies of publications will be deposited before, alongside or after their publication. The consortium will have dedicated part of project's budget to cover the fees associated with open access publication to facilitate dissemination and reuse of the project's results. Project participants, for various reasons, may need to submit articles to journals (or proceedings) that only offer a lower level of open access, requiring either parallel publication or an embargo period. The need for this will be evaluated on a case-by-case basis and will balance benefits against the less convenient or delayed access to the result. In any case, at least the final author's version of every accepted paper will be made publicly available, in accordance with the rules posed by many journals, e.g., for the case of IEEE publications. In this respect, the project will use widely self-archiving (or green open access) services for the research community such as ResearchGate (or Academia) that will allow balancing between traditional publications & open access. Dissemination leading partner VILABS (WP13), with the support of ENoLL and the rest partners will be responsible to track publications and the level of their compliance to open access, as part of dissemination task (T13.1). A record will be kept including all IP assets that the project creates and plans to create. This will be augmented by periodic discussions at the project's management board, reviewing the IP asset list and checking whether exploitation paths are explored for the most valuable IP. Also, for IPR protection, VITALISE will put in place a simple publication notification procedure, so that partners wishing to publish or otherwise disclose project activities and results can notify the other partners in good time. The publication notification and IP asset tracking procedures will be lightweight, requiring partners to fill in simple templates that will be shared via VITILISE's document management tool, considering the classification assigned to the planned deliverables according to the rules of general dissemination focus or restricted to the consortium."*

*"****Open Access Policies:***
Technological developments, development of new ICTs, assisting devices and applications will be rapidly advancing, when possible, through Open Source software projects. At the same time, this will allow to speed up the collaborative technology development across the main Harmonization Board and Body. Additionally, as the aim of VITALISE is to provide easy access to research infrastructures, the consortium will commit to promote new technological advancements within project's context, as open source projects in order to attract the interest of developers outside the

consortium. For this, a dedicated code-sharing repository will be registered to publish matured software releases (e.g. GitLab)"

1.3. Alignment with the principles of fair data handling

All DMPs under the auspices of the Horizon 2020 Work Programme should be in compliance with the FAIR Data Handling Principles (H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, 2016). The FAIR acronym stands for Findable, Accessible, Interoperable, and Reusable, referring to the research outcomes of a project, and most specifically those made available in digital form. The EC deems the FAIR principles to have been fulfilled when a DMP includes the following information:

- "The handling of research data during and after the end of the project"
- "What data will be collected, processed and/or generated"
- "Which methodology and standards will be applied"
- "Whether data will be shared/made open access"
- "How data will be curated and preserved (including after the end of the project)".

Adhering to the FAIR guidelines concerning research data, for each dataset collected, processed, and/or generated during the VITALISE project, the previously mentioned topics are addressed in Section 2 as stated bellow:

- Section 2.1: Datasets, reference and name (details regarding the data collected and generated)
- Section 2.2: Standards and Metadata (describes methodologies and standards that will be implemented)
- Section 2.3: Data sharing, access and preservation (information about sharing, open access, preservation, and archiving of data during and after the project)

1.4. Structure of the document

The deliverable is structured as follows:

- **Section 1** (this section) describes the overall purpose of the document and its target audience. It also provides a summary of the provisions of the VITALISE Grant Agreement and the H2020 provisions for open access to research data. Additionally, it expands upon the adherence of the DMP to the EU H2020 guidelines and the FAIR data handling principles;
- **Section 2** includes an initial overview of the data to be collected, processed and/or generated, followed by a brief presentation of the methodologies and standards to be utilized, and, lastly, it addresses aspects in relation to data security and ethics;
- **Section 3** outlines the allocation of the DMP-related resources;
- **Section 4** provides an overview of the provisions in place for data security;
- **Section 5** presents the methodologies that will be utilized for the anonymization of data;
- **Section 6** addresses the ethical and legal issues that stem from data sharing;
- **Section 7** condenses responds concerning the project's ability to make use of the national/funder/sectional/departmental procedures for data management;
- **Section 8** presents the conclusion and a brief description of the follow-up activities;
- **Section 9** includes the references;
- **Section 10** details the result of the VITALISE data survey and the FAIR forms as completed for the project's each dataset

1.5. Audience

Deliverable D1.4 constitutes a public document; thus it will be made publicly available and released under a Creative Commons License (Creative Commons — CC0 1.0 Universal, n.d.).

2. Data summary

A variety of datasets related to mobility monitoring, biosignals and quality of life (QoL) will be collected for the purposes of the VITALISE project. The partners started defining the type of data that will be collected throughout the project and the reasons and need to collect that data. The Data management team conducted a data survey with the consortium using the DMP template for H2020 projects (Koumoulos et al., 2019). The results of the survey can be found under the annexe section.

This section of the DMP provides details on the datasets to be generated in the VITALISE project as of now. Following version(s) of the DMP may expand on this to include additional datasets and provide a more complete overview of the datasets to be generated, collected and processed in the duration of the project. In accordance with the DMP guidelines by the European Commission the next subsections provide information regarding: a) the description, reference and name of the datasets, b) the standards and metadata that will be used, c) data sharing, access, and preservation, as well as data archiving, in the duration and after the completion of the project.

2.1. Datasets, reference and name

To facilitate the management of data, each dataset was slotted into a distinct category and attributed with a characteristic name, which may also be used as an identifier code for the dataset.

The VITALISE datasets have been initially divided, for the purposes of this deliverable, into five categories:

1. Personal and clinical data
2. Participant Generated Health (PGH) data
3. Output data
4. Requirements' data
5. Other data

The name of each dataset consists of:

1. A prefix **"DS",** denoting the dataset
2. A distinct, unique identification number corresponding to the category and sub-category the dataset may belong to
3. A short descriptive name that provides information regarding the datasets content and use.

For example, the activity captured by a floor sensor constitutes a dataset that belongs to the PG dataset category related to the mobility/activity of the participant and is names as "**DS2.1 – HumanActivityAnalysis**", where "DS" is the prefix, "2.1" denotes the category (Participant Generated data) and subcategory (Participant Activity), and "HumanActivityAnalysis" is the descriptive name.

The name of all the produced data files will consist of the "VIT" prefix, followed by the abbreviated name of the organization responsible with producing this data file, a short descriptive name, a number referring to the timing of the file, and the unique ID denoting the participant to whom the data collected is related to i.e., VIT_Org_Type_Time_SubID.

The datasets are described in greater detail in ANNEX I of the DMP, which also presents the response of the Consortium partners to the FAIR form (Koumoulos et al., 2019). These responses may updated in the following version(s) of the DMP to provide the complete definition of the VITALISE datasets, with special emphasis on their owner and each datasets important characteristics.

Table 1 - initial outline of datasets that may be collected during the vitalise project

#	Dataset Name	Description
	PERSONAL & CLINICAL DATA	

1.1	DS1.1 – ElectronicHealthRecordData	It contains information regarding the participants medical evaluation, i.e., health screening results, medical history etc.
PARTICIPANT GENERATED HEALTH DATA		
2.1	DS2.1 – HumanActivityAnalysis	It contains data related to a participant's activity such as their location, mobility and activities within a room, as derived from sensors (wearable or not), cameras and microphones.
2.2	DS2.2 – SensorHealthData	It contains data related to the participants' health and emotional status as measures by sensors, such as wearable devices (e.g. smartwatches) and activity trackers.
2.3	DS2.3 – SensorEnvironmentalData	It contains data related to the participants' environment as measured by sensors.
2.4	DS2.4 – PROMs	It contains longitudinal Participant Reported Outcome Measures (PROMs) used to monitor and evaluate the participant's functionality and QoL.
OUTPUT		
3.1	DS3.1 – ParticipantStatusOvertime	It contains activity, biosignal monitoring and emotion estimation data, along with questionnaire data, collected from the participants
3.2	DS3.2 – ParticipantHealthLevelPredictiveModel	It contains a combination of a variety of factors that were engineered by extracting them from heterogeneous data sources, and will be used to build predictive models regarding the patients' health
3.3	DS3.3 – SyntheticData	It contains data synthesized from simulation scenarios based on real-life data.
REQUIREMENTS		
4.1	DS4.1 – StakeholderInputData	It contains data collected from focus groups, as well as questions and corresponding answers from the VITALISE Web Survey, conducted in the requirements and system specifications identification phase.
OTHER		
5.1	DS5.1 – CommunicationContactsData	It contains details (e.g., emails) collected from third-parties organization and individuals interested in the VITALISE advancements by subscribing to newsletters, etc.

2.2. Standards and metadata

The initial version of the DMP does not contain a complete listing of all metadata regarding the data that will be produced in the duration of the VITALISE project, even so several domains, which subtend to different rules and directions, that are being considered. Metadata may consist of two types of formats: text-based documents for context-related information, for full description of the data, and JSON according to ISO/IEC 21778 for ISO 19115 standard metadata, for compatibility with international standards.

This is an initial definition of standards:

- Microsoft Office 365 will be used for text-based documents (.doc, .docx, .xls, .xlsx, .ppt, .pptx). Additionally, in cases where larger datasets are concerned, .csv and .txt formats may be used. All finalized documents that have received approval may be provided in .pdf format;
- Illustrations and graphics may utilize Microsoft Vision, with file format .vsd, Photoshop, with different types of file format and most prominently .png, and may finally be available in files with formats .png, .jpg, and .pdf;
- EEG recordings will use the EDF format;
- Audio files may be made available in MP3 and WAV formats;
- Video files may make use of MP4 or Windows Media Video

These specific file formats have been suggested because of their widespread use as accepted standards. Finally, files may be converted, when feasible, to open file formats for long-term preservation.

2.3. Data sharing, access and preservation

VITALISE goal is to support open research data recommendation by the EC. As such, the research data created by the project will be curated depending on the ability to be openly shared and sharing policies. The aim is to preserve the data and keep it accessible and available to the scientific community for the duration of the project and beyond. An Application Programming Interface (API) will be provided to interested users allowing them the access to the platform after registering. At this point, we can assure that at least the following measures will be considered for assuring a proper management of data:

- Dataset minimisation. Only the minimum amount of data will be stored so as to avoid potential risks.
- Access control list for user and data authentication.
- Monitoring and Log of activity.
- Implementation of an alert system that triggers in real time in case of violation of procedures or unauthorized access attempts.
- Liability. A person responsible for the datasets stored will be appointed.
- When possible, the research data will be additionally uploaded to the EC launched Zenodo platform (https://zenodo.org/).

Zenodo is an EU-backed portal based on the well-established GIT version control system (https://git-scm.com) and the Digital Object Identifier (DOI) system (http://www.doi.org). It enables researchers to share and preserve research data and outputs. The repository services offered by Zenodo are free of charge and are meant for a variety of research data: datasets, images, software and more. The platform utilises well-established practices of data preservation such as mirroring and backups. Additionally, each submission of data receives a unique DOI rendering it traceable and citable.

3. Allocation of resources

There are no immediate costs anticipated to make the open results generated in VITALISE FAIR. Additional details will be reported, if needed, in future versions of the DMP. Any unforeseen costs related to open access to research data in Horizon 2020 are eligible for reimbursement during the duration of the project under the conditions defined in the Grant Agreement.

Regarding data management, this will be done as part of WP1 (Project Management). For the time being, ENoLL and the DPO of AUTH lab of Medical Physics and Digital Innovation are responsible for the data management in the VITALISE project.

4. Data security

For the VITALISE project, Data security remains highly relevant and is part of an ongoing process of continuous improvement. VITALISE will take measures to ensure Confidentiality, Integrity and strives for confidentiality by protecting data against unauthorized access, use and distribution. Access controls and user privileges are also considered among the VITALISE consortium.

All generated and collected data will be securely managed, in the duration of the VITALISE project, to protect against unauthorized access and ensure there is no loss or leak of information. Personal data will be made available for access only to authorized personnel. This responsibility concerns all partners, hence all of them will ascertain that all data is protected by complying to all security and access controls within their institution.

The VITALISE security controls constitute the implementation of pseudonymization techniques to eliminate direct identifiers from the data and ensure that subject data cannot be identified without additional information, and the application of encryption (on specific files, and storage devices). They also include the assurance that no unnecessary, unauthorized access, processing, copying and transfer of personal data is carried out and that any processing of such data is performed only for specific purposes of the VITALISE project, through authorized devices.

A part of Deliverable 14.2 of the VITALISE project will include the rules, rights, and restrictions of distinct user-roles regarding the processes and procedures related to personal data. Additionally, it may provide definitions and details regarding the segregation of access control roles (concerning access request, authorization, and administration). This policy may, also, delineate the establishment of an access control system for creating, approving, reviewing, and deleting user accounts, and an authentication mechanism allowing user access to the project's infrastructures.

Integrity is the second area of work for data security. VITALISE looks to ensure the collection of the data is correct, accurate and consistent over the project life cycle. Members in the consortium will need to validate data inputs and data outputs; protect against hardware or system errors; use safeguards to reduce human error among others. In addition, VITALISE ensures the data and data processing is available when they are needed. VITALISE looks to implement Backups for data; Disaster recovery plan; cloud-base solutions; security incidents.

In case of a data breach, members of the consortium are tasked with the immediate notification of the relevant national supervisory authorities and the subjects that the breach of information may impact, as well as documenting all breaches of personal data and any other related information.

5. Data anonymization

The VITALISE anonymization scheme may contain four individual phases: the pre-anonymization, the first layer of anonymization, the elimination of variables, and the final layer of anonymization.

The pre-anonymization marks the initiation of the anonymization process. It may consist of:

- The identification of variables that may be direct (information about an individual that allows their direct identification) or indirect (information that, when combined, may lead to the identification of the individual) identifiers, especially sensitive information (as detailed in GDPR article 9), and other confidential information;
- The classification and scaling of sensitivity of the variables according to their category. For example, direct identifiers (microdata), geographic identifiers, numerical, temporal, metadata (indirect identifiers), especially protected data, etc. The sensitivity scale resulting from the classification is a rudimentary factor of the risk analysis and the Data Protection Impact Assessment (DPIA);

- The determination of variables that cannot be anonymized and thus have to be eliminated prior to anonymization;
- The definition of variables (to be anonymized) that are necessary for achieving the purpose for which the anonymized data will be utilized and avoid any distortion of results with respect to the non-anonymized data;
- The identification of variables that introduce difficulties in the anonymization process;
- The establishment of the protection policy to ensure people's privacy, after variable categorization, and limit the amount of sensitive information used in the anonymization.

The first layer of anonymization is, essentially, the pseudo-anonymization of datasets to disassociate them from their direct identifiers, a necessary step for data security as well.

The third phase would be the elimination of variables that cannot be anonymized and may contain direct identifiers for specific subjects or groups. The elimination of these variables may be achieved through the aggregation of its range to hide the individual. For example, in the case of tiny groups, their information can be masked by mixing their data with a group with a larger range (adding, if necessary, a percentile reference to mark the existence of the smaller group within a larger one).

The final phase consists of irreversibly disassociating identifiers from the datasets. In this phase, variables categorized to contain indirect identifiers (e.g., data of birth, with exact date) are subjected to aggregation through k-anonymization so that this information may not be wholly accessible during the analysis, but the quality of the results is maintained. If the implementation of this process is not possible other masking techniques will be utilized to ensure that no person can be re-identified through this information.

Regarding special types of data

During anonymization, certain variables, such as biometric data, voice or image recording, introduce difficulties, which should be dealt with in the first phases of the process.

Regarding voice recordings, the task of anonymizing may not be always feasible depending on the purposes of processing. For example, if the voice recording is to be analyzed for emotion detection, anonymization will effectively render the recordings unusable. If the purposes of collection allow it, and preservation is necessary for the purposes of the project, transcribing the related elimination of possible identifiers (autochthonous expressions, epideictic elements, rhetorical identifiers, etc.) allows for the later reproduction of this transcript through a synthesizing voice device, thus anonymizing it. Alternatively, dedicated software for anonymizing will be used (e.g. VoiceMask, Voice Transformer) (Qian et al., 2019; Voice Transformer H2020 COMPRISE Project, 2020).

Images always carry the risk of the re-identifying the subject (through the subject's environment instead of their genetic features) within them and as such, in some cases, the image may require special processing to prevent this issue. In addition to face blurring or pixilation, in instances where the image contains a tattoo or a scar on a person's skin that may reveal their identity, the image must be subjected to specific digital processing to make the re-identification of the individual impossible.

Concerning biometric data, the purpose of this information may present limitations to the extent of its anonymization to avoid any distortion in regard to the non-anonymized data. These limitations should be considered in the initial stages of anonymization, especially in the DPIA, as a risk to the irreversibility of the process. In this instance, the implementation of encryption techniques for restricted and controlled access to biometric information should be clearly articulated in the DPIA itself.

6. Ethical aspects

VITALISE partners have to comply with the ethical principles as set out in Article 34 Ethics and Research Integrity of the Grant Agreement, which state that all activities must be carried out in compliance with:

(a) ethical principles (including the highest standards of research integrity)

and

(b) applicable international, EU and national law.

The content above has been included in the VITALISE DoA(Grant Agreement Number 101007990 - VITALISE, 2021) and will be taken into consideration for the Data Management Plan. In addition, Task 1.4 Ethical and Security Management will work on 2 reports on ethical and safety issues along the project. Sensor, tools and devices are being collected in Task 2.1 and Task 1.4 for an evaluation under the Ethics principles.

Task 1.4 started working together with task 1.5 Data management to ensure the development of an appropriate ethics code of conduct of research for VITALISE.

The code of conduct plans to include consent forms, access rights of patient data, use of patient's medical and personal data in project deliverables and publications, etc.

6.1. Rights of data subjects

6.1.1. The right to be informed

According to GDPR art. 13, where personal data relating to the data subjects are to be processed, they will be provided with a) the identity and contact details of the controller, b) the contact details of the DPO, c) the purpose and legal basis of processing, d) the recipients of the personal data and e) where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of adequacy decision by the Commission. In addition, the controller with inform the data subjects about a) the period of retainment, b) the existence of rights of the data subject, and c) the right to lodge a complaint to a supervisory authority.

6.1.2. The right of access

Specific mechanisms will be provided to the data subjects to access their personal data and supplementary information, allowing them to be aware of and verify the lawfulness of the processing.

6.1.3. The right to rectification

Data subjects will have an ability to correct data sets, when personal data are found to be inaccurate.

6.1.4. The right to erasure

Data subjects will have an ability to issue a request for all data erasure and, within predefined but non-later that within 40 days after request, VITALISE administrator will delete all corresponding to data Subject data sets.

6.1.5. The right to restrict processing

With additional stipulations, data subjects will have the ability to limit the processing of their personal data, as defined during consent processing.

6.1.6. The right to data portability

Data subjects will have an ability to issue a request for all data concerning them, within 40 days after request.

6.1.7. The right to object

Data subjects will have the ability to object at any time to processing of their personal data. Processing of data sets corresponding to them will be set on hold and excluded from processing till consent on processing is given back.

6.1.8. The right regarding automated decision-making

VITALISE will not base a decision solely on automated means, including profiling, which produces legal or similar effects.

6.1.9. Other rights

Data subjects participating in research activities reserve the right to withdraw their consent and notification in case of data breach.

7. Other issues

Currently, the VITALISE project has not made provisions, regarding data management, for the use of other process and techniques other than the ones presented in this initial version of the DMP.

8. Conclusions and future work

This first version of the DMP outlines the data management policy that will be applied in the premises of the VITALISE project, in line with the open data principles and FAIR guidelines data management of Horizon 2020.

The FAIR principles provide a comprehensive and straightforward guide to correctly implementing procedures of administrative and technical nature, as well as underline the responsibilities for the unimpeded integration of data management in the VITALISE project. These procedures are implemented through the use of state-of-the-art tools and standards, such as the OpenAIRE initiative and the Zenodo depository for research data, to guarantee that the VITALISE results (open data, open science publications, and open-source software included) are properly preserved and remain accessible and available for use even after the end of the project's lifecycle.

The current version of the document details the consolidated feedback form all partners, as well as their provisions for datasets in whose collection and generation they may be involved. So far, for the purposes of the project there have been identified ten (10) distinct datasets that will be produced/collected in the duration of the VITALISE project. Each dataset is accompanied by a questionnaire related to the FAIR principles.

The VITALISE DMP, is a document that will continue to be updated regularly (at least once per year) to include new results extracted from the ongoing data survey, or to update any changes in the common data management policy. Moreover, data management is an ongoing task that is the concern of almost all VITALISE WPs.

9. References

Creative Commons — CC0 1.0 Universal. (n.d.). Retrieved September 21, 2021, from https://creativecommons.org/publicdomain/zero/1.0/

General Data Protection Regulation (EU GDPR) 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 Di. (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e40-1-1

Grant Agreement Number 101007990 - VITALISE. (n.d.).

H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. (2016).

Koumoulos, E. P., Sebastiani, M., Romanos, N., Kalogerini, M., & Charitidis, C. (2019). *Data Management Plan template for H2020 projects*. https://doi.org/10.5281/ZENODO.2635768

Open access - H2020 Online Manual. (2017). https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

OpenAIRE. (2017). *What is the Open Research Data Pilot*. https://www.openaire.eu/what-is-the-open-research-data-pilot

Qian, J., Du, H., Hou, J., Chen, L., Jung, T., & Li, X. (2019). Speech Sanitizer: Speech Content Desensitization and Voice Anonymization. *IEEE Transactions on Dependable and Secure Computing*. https://doi.org/10.1109/TDSC.2019.2960239

Voice Transformer H2020 COMPRISE project. (2020). https://gitlab.inria.fr/comprise/voice_transformation

10. ANNEX I – Vitalise data survey

10.1. Personal and clinical data

10.1.1. DS1.1 – ElectronicHealthRecordData

DATASET REFERENCE NAME		**DS1.1 – ElectronicHealthRecordData**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	The purpose of collected data is to generate different indexes and indicators from the health records of participants for capturing and assessing their status	1. Identification of most important information to support the transition decision.
			2. Create recommendations that can support healthcare professionals.
	Type and Format of Data	**Form**	**Format**
	Text	EHR data, such as diagnoses, and treatment	.json, .doc
	Numeric	EHR data, such as diagnoses, and treatment	.json, .xlsx, .csv
	Audiovisual	-	-
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
		-	-
	Reused-Data (rd)	no	no reuse of existing data
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	

	Observational	Original data, captured from the patients EHR	
	Experimental	-	
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Fixed	
	Quantity	<1	*in MB/GB*
		-	-
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	Excel tables	CSV data are interlinked with corresponding XML documents.
	Identifiability of data (refer to standard id mechanisms)	Data will be assigned a unique ID, referring to data subject, data type and time of collection	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	
	Search keywords approach	Probably yes, index key numbers corresponding to each simulated dataset.	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	The dataset will be accompanied with detailed documentation of its content.	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data produced in their entirety belong to partners and are not made openly available.	

	Data kept closed	VITALISE partners only	Confidential PII data that require ethical permit for use
	How data will be made available	-	
	Methods or software (SW) tools for data access	Standard text reader/editor	
	SW documentation and other information needed	N/A	
	Repository for deposit of data, metadata, documentation and code	*Private*	*Private*
	Access restrictions	Data is restricted to VITALISE partners	During the duration of the project, data are stored up to 10 years from collection according to law
	Data interoperability assessment	N/A -	N/A
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	Plain text and ICD-10 classification where applicable.	
	Data licensing for wide reuse	N/A	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	-	TBD
	Data usability by Third Parties (after the end of the project)	N/A	
	Restrictions to data re-use	TBD	TBD

	Quality assurance process	Regular back up	Data entry, access and processing is registered, and password protected.
	Length of time of data reusability	Within the timeframe of the project	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers.	

10.2. Patient generated health data

10.2.1. DS2.1 – HumanActivityAnalysis

DATASET REFERENCE NAME	**DS2.1 – HumanActivityAnalysis**	
1. DATA SUMMARY		
Purpose of the Data	**Purpose**	**Objectives**
Purpose	1. Location and activity of participants in Living Labs 2. Participant activity at home.	1. Creation of skeleton-joints datasets
		2. Create recommendations that can support healthcare professionals.
		3. Creation of dynamic personas for each participant, for personalizing and improving care intervention after a rehabilitation period
Type and Format of Data	**Form**	**Format**
Text	-	-
	-	-
Numeric	Floor, bed, proximity sensor	.json
	Movement	.json
	Contact pressure	.json
Audiovisual	3D depth sensor	.mp4
	RGB camera	.mp4
Simulated		
model	-	
model type	-	
computer code	-	
data type	-	

	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
		Sensorial Gradior: Virtual reality headset (Quest 2) *Active Gradior*: Kinect v2	*.json*
	Reused-Data (rd)	no	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Data captured in real time from Real-Time Location System (RTLS) and Home area location, sensor readings, activity tracking, live recordings, interaction flows, logs.	
	Experimental	Joint measurement of the upper extremities using the Captiv L7000 measurement system (TEA ERGO)	
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	
	Quantity	10 GB x month x user	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-
	Search keywords approach	-	
	Clear versioning	**Versioning**	**Traceability**

	approach	-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Due to the sensitive nature of data and for ethical reasons, the dataset will be restricted to public access. Some data should never be shared, while some of it could be accessed externally upon request for research purposes
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data access	TBD	
	SW documentation and other information needed	TBD	
	Repository for deposit of data, metadata, documentation and code	TBD	
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			

	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.2.2. DS2.2 – SensorHealthData

DATASET REFERENCE NAME		**DS2.2 – SensorHealthData**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	Activity level, biosignals monitoring and emotion estimation in clinical, Living Lab, and home environment.	1. Identify range of improvement in mobility 2. Create recommendations that can support healthcare professionals. 3. Creation of dynamic personas for each participant, for personalizing and improving care intervention after a rehabilitation period
	Type and Format of Data	**Form**	**Format**

	Text	-	-
		-	-
	Numeric	HR, HRV, BP, GSR, SPO2	.json, .csv
		EEG, EMG, ECG, EOG	.edf, .csv
	Audiovisual	Microphone	.mp3 or .wav
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
	Reused-Data (rd)	no	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Data captured in real time from sensors, wearables and microphone.	
	Experimental		
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	
	Quantity	1 GB x month x user	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			

	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-
	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Due to the sensitive nature of data and for ethical reasons, the dataset will be restricted to public access. Some data should never be shared, while some of it could be accessed externally upon request for research purposes
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data access	TBD	
	SW documentation and other information needed	TBD	
	Repository for deposit of data, metadata, documentation and code	TBD	

	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.2.3. DS2.3 – SensorEnvironmentData

DATASET REFERENCE NAME	**DS2.3 – SensorEnvironmentData**
1. DATA SUMMARY	

	Purpose of the Data	Purpose	Objectives
	Purpose	Activity tracking of participants based on interactions with their environment in Living Labs, and home.	1. Identify range of improvement in mobility
			2. Create recommendations that can support healthcare professionals.
			3. Creation of dynamic personas for each participant, for personalizing and improving care intervention after a rehabilitation period
	Type and Format of Data	**Form**	**Format**
	Text	-	-
		-	-
	Numeric	Window and Door sensor Temperature sensor Flood sensors Luminosity sensor Air quality sensor	.json, .csv
	Audiovisual		
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
	Reused-Data (rd)	no	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Data captured in real time regarding the participants' environment from sensor readings, interactions, logs.	
	Experimental		
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	

	Quantity	1 GB x month x user	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-
	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Due to the sensitive nature of data and for ethical reasons, the dataset will be restricted to public access. Some data should never be shared, while some of it could be accessed externally upon request for research purposes

	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data access	TBD	
	SW documentation and other information needed	TBD	
	Repository for deposit of data, metadata, documentation and code	TBD	
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data	None	

	reusability		
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.2.4. DS2.4 – PROMs

DATASET REFERENCE NAME		**DS2.4 – PROMS**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	To collect data regarding the participants cognitive and physical health, and habits.	1. Identification of most important information to support the transition decision.
			2. Create recommendations that can support healthcare professionals.
			3. Creation of dynamic personas for each participant, for personalizing and improving care intervention after a rehabilitation period
			4. Assess the effectiveness and impact of those physical and cognitive interventions around healthy spaces, sports and performing arts
	Type and Format of Data	**Form**	**Format**
	Text	PROM data form questionnaires	.json, .xlsx,, .csv
		-	-
	Numeric	PROM data from questionnaire scales	.json .xlsx, .csv
		-	-
	Audiovisual	-	-
	Simulated		
	model	-	
	model type	-	

	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		N/A	-
	Instrument specific	*equipment*	*format*
		N/A	-
	Reused-Data (rd)	*No*	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Original data, captured through 1) field test measurements carried out in real time, and 2) interactive questionnaires/scales	
	Experimental	-	
	Simulation	-	
	Derived/Compiled	-	
	Reference or Canonical (links)	-	-
	Dataset is:	**Growing**	
	Quantity	<1	*in MB/GB*
		-	-
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-
	Search keywords approach	TBD	
	Clear versioning	**Versioning**	**Traceability**

	approach	-	-
		-	-
	Standards or procedures for metadata creation applied	The dataset will be accompanied with detailed documentation of its content.	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Confidential data that require ethical permit for use.
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data access	-	-
	SW documentation and other information needed	-	
	Repository for deposit of data, metadata, documentation and code	TBD	-
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	Will depend on publisher policy
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	

	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	TBD
	Quality assurance process	Regular back up	TBD
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.3. Output

10.3.1. DS3.1 – ParticipantStatusOvertime

DATASET REFERENCE NAME		**DS3.1 – ParticipantStatusOvertime**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	To identify shifts and trends in the participants' health status, in relation to functionality, mobility, QoL, etc., overtime. The dataset includes daily life observations and measurements gathered from clinical, LivingLab, and the participants' home environment.	1. Identify range of improvement in mobility
			2. Create recommendations that can support healthcare professionals.
			3. Creation of dynamic personas for each participant, for personalizing and improving care intervention after a rehabilitation period
			4. Assess the effectiveness and impact of those physical and cognitive interventions around healthy spaces, sports and performing arts
	Type and Format of Data	**Form**	**Format**
	Text	-	-
		-	-
	Numeric	Health status Indeces	.json, .xls

		Standardized clinical scales	.json, .xls
	Audiovisual	-	-
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
	Reused-Data (rd)	no	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Data captured in real time from sensors, wearables, cameras and microphones in clinical, LivingLab and home (real-life scenarios (in the wild)) environments.	
	Experimental		
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	
	Quantity	16GB	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the	-

		VIT_Org_Type_Time_SubjID format.	
	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Due to the sensitive nature of data and for ethical reasons, the dataset will be restricted to public access. Some data should never be shared, while some of it could be accessed externally upon request for research purposes
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data access	TBD	
	SW documentation and other information needed	TBD	
	Repository for deposit of data, metadata, documentation and code	TBD	
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			

	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.3.2. DS3.2 – ParticipantHealthLevelPredictiveModel

DATASET REFERENCE NAME		**DS3.2 – ParticipantHealthLevelPredictiveModel**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**

	Purpose	Support the creation of the predictive models from WP5-7	To predict the course of the participants' health, based on the data collected form a clinical, Living Lab, and home environment.
	Type and Format of Data	**Form**	**Format**
	Text	Clinical data merged from different sources	.json, .csv
		New features derived from existing data	.json, .csv
	Numeric	Clinical data merged from different sources	.json, .csv
		New features derived from existing data	.json, .csv
	Audiovisual	-	-
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
	Reused-Data (rd)	no	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Data captured in real time from sensors, wearables, cameras and microphones in clinical, LivingLab and home (real-life scenarios (in the wild)) environments.	
	Experimental		
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	
	Quantity	16GB	in MB/GB

	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-
	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Due to the sensitive nature of data and for ethical reasons, the dataset will be restricted to public access. Some data should never be shared, while some of it could be accessed externally upon request for research purposes
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	

	Methods or software (SW) tools for data access	TBD	
	SW documentation and other information needed	TBD	
	Repository for deposit of data, metadata, documentation and code	TBD	
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			

	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.3.3. DS3.3 – SyntheticData

DATASET REFERENCE NAME		**DS3.3 – SyntheticData**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	To create synthetic data for each dataset available	Protect the participants' privacy when sharing data.
	Type and Format of Data	**Form**	**Format**
	Text	PROM data form questionnaires	.json, .xlsx, .csv
			.
	Numeric	PROM data form questionnaires	.json, .xlsx, .csv
		Floor, bed, proximity sensor Movement Contact pressure	.json
		HR, HRV, BP, GSR, SPO2	.json, .csv
		EEG, EMG, ECG, EOG	.edf and .csv
		Window and Door sensor Temperature sensor Flood sensors Luminosity sensor Air quality sensor	.json, .csv
	Audiovisual	3D depth sensor	.mp4

	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
		Sensorial Gradior: Virtual reality headset (Quest 2) *Active Gradior*: Kinect v2	*.json*
	Reused-Data (rd)	no	-
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational		
	Experimental		
	Simulation	Data generated using Generative Adversarial Networks	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	
	Quantity	TBD	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-

	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	For research publication purposes.
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data access	TBD	
	SW documentation and other information needed	TBD	
	Repository for deposit of data, metadata, documentation and code	TBD	
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			

	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.4. Requirements

10.4.1. DS4.1 – StakeholderInputData

DATASET REFERENCE NAME		**DS4.1 – StakeholderInputData**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	To receive feedback and input from the participants, family members, healthcare professionals and indirect caretaker, regarding the intervention schemes. To explore (a) the perception of end-users/participants about their health needs, priority outcomes and care requirements, and (b) the participants' view, preferences and	To identify issues causing low adherence in rehabilitation programs using technology, propose changes to existing intervention tools, and identify the most important information to support the transition decision Identify/establish the health needs, priority outcomes, and care requirements of VITALISE participant, as well as their views, preferences and expectations.

		expectations regarding the JRAs.	
	Type and Format of Data	**Form**	**Format**
	Text	Interviews aggregated responses	.json, .txt, .csv, .xlsx
		Survey and interview aggregated data from content analysis	.json, .xlsx, .csv
	Numeric	-	-
		-	-
	Audiovisual	-	-
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
		-	-
	Reused-Data (rd)	no	*Existing data could NEVER be used for other research projects*
	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Original data captured through (a) interactive activities (workshop/Delphi) with health care professionals, participants, family member, an informal caretakers (b) online/telephone interviews (one to one or focus group where possible) audio-recorded via use of standard commercial audio recorders, and (b) online surveys.	
	Experimental		
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-

	Dataset is:	Growing	
	Quantity	TBD	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	To all partners of the VITALISE consortium involved in WP5-7.	
		TBD	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	TBD	-
	Identifiability of data (refer to standard id mechanisms)	TBD	-
	Naming conventions used	Folder names will derive from the data type. Files will be named based on the VIT_Org_Type_Time_SubjID format.	-
	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		-	-
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			
	Data openly available	Data will be made available upon request	
	Data kept closed	Restricted public access	Confidential data that require ethical permit for use.
	How data will be made available	After being anonymized, data can be uploaded to a shared repository	
	Methods or software (SW) tools for data	-	

	access		
	SW documentation and other information needed	-	
	Repository for deposit of data, metadata, documentation and code	TBD	
	Access restrictions	TBD	TBD
	Data interoperability assessment	-	-
		-	
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	TBD	
	Data licensing for wide reuse	Open Access	Creative Commons Attribution 4.0 International (CC-BY)
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	Two-three (2-3) years after the project end.	-
	Data usability by Third Parties (after the end of the project)	Third parties may access the dataset after request	
	Restrictions to data re-use	Reference to publications from which datasets have been produced are required	To acknowledge properly the authors' work
	Quality assurance process	Regular back up	Data entry, access and processing in general is password protected and registered. Independent in-house researcher validates the data.
	Length of time of data reusability	None	
3 ALLOCATION OF RESOURCES			

	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	-
	Data Management Responsibilities	Living lab VITALISE node responsible, with the support of infrastructure tools deployment lead and technical developers	

10.5. Other

10.5.1. DS5.1 - CommunicationContactsDataset

DATASET REFERENCE NAME		**DS5.1 – CommunicationContactsDataset**	
1. DATA SUMMARY			
	Purpose of the Data	**Purpose**	**Objectives**
	Purpose	To store contact details (e.g., emails) gathered from third parties, organizations and individuals interested in VITALISE advancements by subscribing to newsletters, etc.	To implement a rigorous communication strategy.
	Type and Format of Data	**Form**	**Format**
	Text	Contact details (e.g., emails)	.json, .txt, .csv, .xlsx
	Numeric	-	-
		-	-
	Audiovisual	-	-
	Simulated		
	model	-	
	model type	-	
	computer code	-	
	data type	-	
	format	-	
	Discipline specific information	*discipline*	*format*
		-	-
	Instrument specific	*equipment*	*format*
		-	-
	Reused-Data (rd)	no	*Existing data could NEVER be used for other research projects*

	Data Origin	***Define & describe the origin/ source of your data. Data can be gathered from different sources.***	
	Observational	Data gathered from third parties, organizations and individuals interested in VITLISE advancements by subscribing to newsletters.	
	Experimental		
	Simulation	-	
	Derived/Compiled	**-**	
	Reference or Canonical (links)	-	-
	Dataset is:	Growing	
	Quantity	*Max 0.5 MB*	in MB/GB
	Data Security & Storage	Data will be stored in secured computer/ hard drive at pilot partners premises	Data is pseudonymized and password protected. Encryption key is stored securely separate from the dataset.
	Data Value (long term)	Digital Marketing Companies.	
		Business analytics and intelligence.	
2 FAIR DATA			
2.1 FAIR DATA - Making data findable			
	Discoverability of data (metadata provision)	-	-
	Identifiability of data (refer to standard id mechanisms)	-	-
	Naming conventions used	-.	-
	Search keywords approach	-	
	Clear versioning approach	**Versioning**	**Traceability**
		date of extraction	date of extraction
		-	-
	Standards or procedures for metadata creation applied	-	
2.2 FAIR DATA -Making data openly accessible			

	Data openly available	Data produced in their entirety belong to partners and are not made openly available.	
	Data kept closed	VITALISE partners only	VITALISE partners only
	How data will be made available	-	
	Methods or software (SW) tools for data access	Standard text reader/editor	Standard text reader/editor
	SW documentation and other information needed	N/A	
	Repository for deposit of data, metadata, documentation and code	*Private*	*Private*
	Access restrictions	Data is restricted to VITALISE partners	Data is restricted to VITALISE partners
	Data interoperability assessment	N/A -	N/A
2.3 FAIR DATA – Making data interoperable			
	Standard vocabulary or mapping to commonly used ontologies	Plain text and ICD-10 classification where applicable.	
	Data licensing for wide reuse	N/A	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)			
	Timing of data availability for re-use (incl. indications on embargo)	-	-
	Data usability by Third Parties (after the end of the project)	N/A	
	Restrictions to data re-use	TBD	TBD
	Quality assurance process	Regular back up	Regular back up

	Length of time of data reusability	Within the timeframe of the project	
3 ALLOCATION OF RESOURCES			
	Costs estimates for making data FAIR	No additional costs to produce and manage data according to FAIR	No additional costs to produce and manage data according to FAIR
	Data Management Responsibilities	Dissemination and exploitation task falling under the responsibility of WP13, leader VILABS	