SustAInLivWork

Project Acronym	SustAInLivWork
Project full title	Centre of Excellence of AI for Sustainable Living and Working
Call identifier:	HORIZON-WIDERA-2022-ACCESS-01-two-stage
Type of Action:	HORIZON-CSA
Start date:	01/09/2023
End date:	31/08/2029
Grant Agreement No.	101059903

D6.2: Data Management Plan (DMP)

Work Package (WP):	WP6: Project Management and Coordination
Document No.:	D6.2
Due date of Deliverable:	29/02/2024
Lead Beneficiary:	Lithuanian University of Health Sciences (LSMU)
Authors:	Dovydas Verikas (LSMU), Arnas Karužas (LSMU), Agnė Dorelaitienė, (LSMU), Gintarė Kučinskienė (LSMU).
Reviewers:	Vilma Karoblienė (KTU), Renata Višinskaitė (KTU), Agnė Paulauskaitė Tarasevičienė (KTU), Oksana Palekienė (KTU), Jurgita Raščiuvienė (VMU), Dovilė Kuizinienė (VMU), Jukka Yrjänäinen (TAU), Alexander Schlaefer (TUHH), Sarah Latus (TUHH).
Type:	DMP
Document dissemination level:	Public

Co-funded by the European Union

List of Project Beneficiaries:

1. Kauno technologijos universitetas (KTU), Coordinator;
2. Vilniaus Gedimino Technikos Universitetas (VILNIUS TECH);
3. Vytauto Didziojo Universitetas (VMU);
4. Lietuvos Sveikatos Mokslu Universitetas (LSMU);
5. Tampere University (TAU);
6. Hamburg University of Technology (TUHH).

Revision history

No.	Authors	Description	Date
1	Agnė Dorelaitienė, (LSMU), Dovydas Verikas (LSMU), Arnas Karužas (LSMU).	First draft (v 0.1)	13/02/2024
2	Dovydas Verikas (LSMU), Arnas Karužas (LSMU), Agnė Dorelaitienė, (LSMU), Gintarė Kučinskienė (LSMU).	Second draft (v 0.2)	26/02/2024
3	Dovydas Verikas (LSMU), Arnas Karužas (LSMU), Agnė Dorelaitienė, (LSMU).	Third draft (v 0.3)	28/02/2024

TABLE OF CONTENT

1. INTRODUCTION 5
1.1 Deliverable description 6
1.2 Project mission and objectives 7
2. DATA SUMMARY 7
2.1 Data reuse 8
2.2 Data collection 9
2.3 Characteristics of AI-related research data 10
2.4 Data formats 11
3. DATA MANAGEMENT GUIDELINES 12
4 FAIR DATA 12
4.1 Making data findable, including provisions for metadata 13
4.2 Making data accessible 13
4.3 Making data interoperable 14
4.4 Increase data re-use 14
5 ALLOCATION OF RESOURCES 14
6 DATA SECURITY 15
7 ETHICS 15
8 REFERENCES 17

List of figures

1. Research data life cycle. 5
2. Project data categories are distinguished. 8

List of tables

1. Examples of data formats 11

GLOSSARY

CA	Consortium Agreement
CoE	Centre of Excellence
GA	Grant Agreement
EC	European Commission
WP	Work Package
AI	Artificial Intelligence
DMP	Data Management Plan
IP	Intellectual Property
ID	Identification Data
GDPR	EU General Data Protection Regulation
PC	Project Coordinator
FAIR	Findable, Accessible, Interoperable, and Reusable
PT	Project Team
HR	Human Resources
NGOs	Non-governmental organizations
UUID	Universally Unique Identifier
DOI	Digital Object Identifier

1 INTRODUCTION

The deliverable Data Management Plan (from now on –DMP) sets the framework for managing data in the project SustAInLivWork along the complete data lifecycle: from data acquisition to curation, use and dissemination as well as beyond the project's lifetime. It describes the data that will be collected, generated, stored, and backed up during the project and explains how it will be exploited or if it will be shared for verification and re-use. Moreover, the DMP details who owns and is responsible for the different data which data is confidential, and which will be shared openly.

SustAInLivWork employs a data lifecycle management approach to ensure that the large pool of data that will be collected over the six-year lifetime of the project follows a consistent, well-established, and standardized approach. Only a traceable workflow throughout the lifetime of data resources can assure good data quality, interoperability, discoverability, and replicability. Beyond the lifecycle perspective on data, the DMP presents how the project plans to make data Findable, Accessible, Interoperable, and Reusable (FAIR).

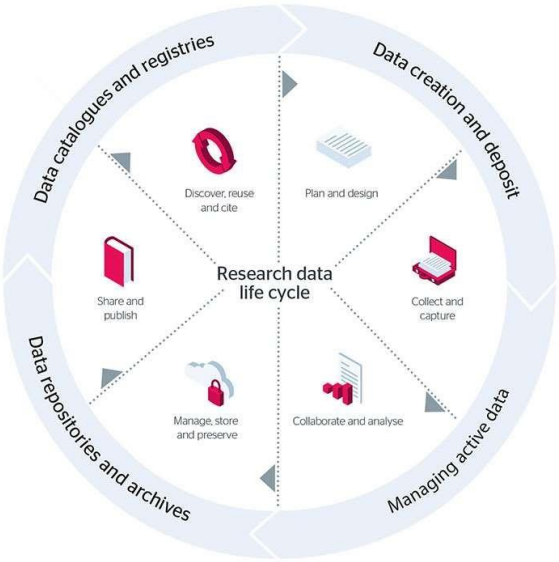

Figure 1. *Research data life cycle*

Reference on source: Medeiros, C.B. (2018). Scientific Data Management – from collection to preservation (Scientific Data Management – from collection to preservation | SciELO in Perspective)

The plan and design phase of the data creation and deposit includes writing and updating the DMP, addressing data and ethics protections as well as relevant IP and copyright issues.

Moreover, decisions on collection methods and file formats will be part of this phase. The collect and capture step includes a collection of data through different methods including AI, data cleaning, and preparation for further processing or analysis. It includes data annotation to ensure traceability and discoverability. Collaboration and analysis will include data preparation, as well as processing, analysis, and/or evaluation.

Secure storage of data in state-of-the-art repositories, procedures for back-up and disaster recovery management as well as dealing with data according to the different sensitivity and privacy levels fall within the manage, store, and preserve step. Depending on the classification of data, those that can be openly shared will either be published in scientific, white, or practice papers or be made available as data sets in public repositories, enabling the discovery, reuse, and citation.

The DMP is a living document that will evolve as project work progresses. The SustAInLivWork

beneficiaries update the DMP every time this is necessary, including but not limited to when new data is required, or upon any relevant changes in the project. The DMP is developed within M6 and will be updated providing periodic reports (within M18, M36, M54, M72).

The DMP gives specific instructions to SustAInLivWork beneficiaries on what is requested from them in specific situations working with different data. Compliance with the DMP is obligatory for all SustAInLivWork beneficiaries.

1.1 Deliverable description

DMP is one of the WP6 "Project Coordination and Management" deliverables. The goal of the WP6 is to develop an effective, transparent, and comprehensive, administrative, financial, and legal management to ensure the successful execution of the SustAInLivWork project.

WP6 consists of six Tasks, that last the whole duration of the project (M1-M72).

- Task 6.1 Project management and coordination
- Task 6.2 Project Data Management
- Task 6.3 Monitoring of project work progress and reporting
- Task 6.4 Project Financial Management and Supervision
- Task 6.5 Project Innovation, Knowledge and IPR Strategy
- Task 6.6 Project Quality Assurance, Risks and Ethics Management

The DMP is output of the Task 6.2.

The DMP is structured into several sections explaining the following key aspects related to the overall project management process in SustAInLivWork project. The document is a DMP for the SustAInLivWork project, a 6-year initiative to create a CoE for AI for Sustainable Living and Working. It outlines the project background, objectives, data types to be collected, data formats, plans to make the data FAIR (Findable, Accessible, Interoperable, Reusable), allocation of resources, data security provisions, and ethical considerations. Both new and existing data will be collected through surveys, workshops, secondary sources in areas like research, software, dissemination materials etc. Making data FAIR is discussed by using unique IDs, rich metadata, data repositories, licensing for reuse. Resources are allocated especially for data management activities. Data security and privacy principles align with EU regulations like EU General Data Protection Regulation (EU GDPR). Informed consent procedures are outlined as an ethical measure. Overall, it is a living document covering data management best practices over the project lifecycle, ensuring traceability, quality, and sharing of project data assets.

The DMP is related to all activities, tasks, and WPs in SustAInLivWork project as DMP will be the basis for the data management rules and procedures to be followed in the execution of SustAInLivWork project to ensure achieving the expected outputs according to project objectives and resources available.

Effective data stewardship is imperative as AI systems rely increasingly on training data's size, diversity, and accuracy. The project implements rigorous data curation protocols maximizing quality and compatibility of accumulated datasets for AI modelling. Standardized management procedures enable the consolidation of vast annotated data resources - fuel for developing advanced AI able to tackle real-world complexity.

1.2 Project mission and objectives

The SustAInLivWork project aims to create a joint CoE of Artificial Intelligence (hereinafter – AI) for Sustainable Living and Working in Lithuania, which would act as a lighthouse, making a significant impact on strengthening the research and innovation ecosystem nationally and internationally. Specialized focus is directed in the development and application of R&I solutions based on AI in the manufacturing, energy, health and transport sectors.

The project objectives are:

1. To create a long-term and independent CoE of AI generating important and sustainable solutions for the region.
2. Conduct research, train, and educate members of society about the benefits and opportunities of AI solutions.
3. Become a driving force for change, an international centre of AI solutions for sustainable living and working.
4. To create a Lithuanian AI cluster that would ensure an effective transformation of sustainability, paying special attention to S3 priority areas.

In order to reach the overall objective of the SustAInLivWork project it will be implemented within 3 main phases – launching, growth and self-sustainability.

SustAInLivWork will create and ensure the long-term self-sustainability of an internationally recognized CoE of AI for sustainable living and working in the Lithuanian region and the whole Baltic Sea Region, that will conduct cutting edge scientific research related to AI in four key sectors: manufacturing, energy, health and transport; that are aligned with the S3 strategy of the country and the SDGs and that will create a critical mass of researchers working on AI both in public and private sectors.

2 DATA SUMMARY

Data collection in SustAInLivWork strives to support the production of freely (re)producible material to create a common computing and data infrastructure between the widening and advanced countries, which is integrable in CoE with a long-lasting mindset ensuring long-term self-sustainability.

The categorization of project data into three distinct categories - **Project management documents, Educational skills related materials,** and **Scientific data** - plays a pivotal role in the effective management and utilization of information (Fig. 2). This division is not merely a structural organization but serves as a strategic framework that acknowledges the unique characteristics and purposes associated with each category:

1. **Project management documents** - highlights the administrative and procedural records essential for the project's governance and compliance. This involves financial data, IP documents, strategies, plans, agendas, HR data, reports (implementation), meeting minutes and other official paperwork that guide the project's trajectory. The nature of formal documents necessitates controlled dissemination to ensure compliance with regulatory standards and to safeguard sensitive information.
2. **Educational skills related materials** - recognizes the distinct nature of content intended for educational purposes. This can encompass a wide range of materials, including courses and trainings, educational programs, teaching material on AI & sustainability topics, technical support documents, etc. Given the varied audience and intended use, the dissemination requirements for educational material differ from those of scientific data. Accessibility and comprehensibility become key considerations, leading to the adoption of user-friendly formats to facilitate broader dissemination and utilization.

3. **Scientific data** - underscores the significance of empirical findings, experimental results, and raw data generated through scientific methodologies among the beneficiaries. This category involves data that holds critical implications for the project's research objectives, necessitating specific considerations in terms of dissemination, storage, and accessibility. For instance, scientific data may include AI-relevant Platform data (DOI persistent data, data sets, metadata & documentation, etc.) and AI-relevant data stored in SustAInLivWork infrastructure (raw data not fully prepared to be stored in Platform).

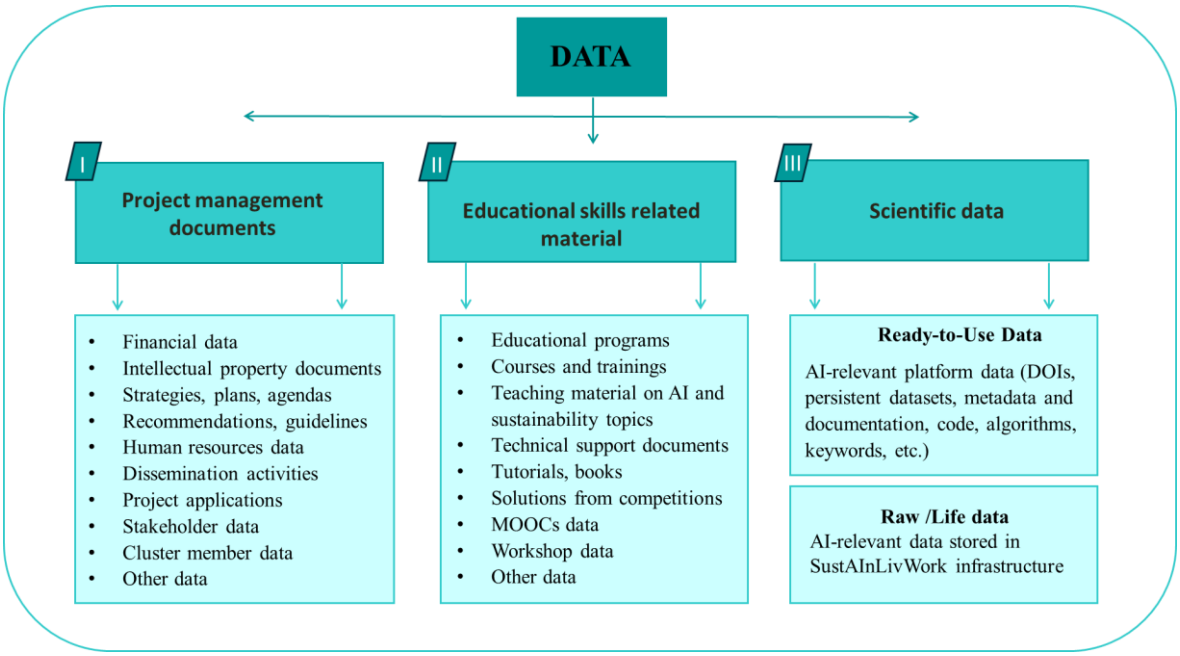

Figure 2. *Project data categories are distinguished*

The next sub-chapters present the main opportunities for data collection and sharing over the project lifetime.

2.1 Data reuse

In the SustAInLivWork project, data reuse is a fundamental aspect to leverage existing knowledge and resources effectively. The project will strategically incorporate various existing datasets, reports, and documents to enhance its research and contribute to the project's objectives. The following outlines the planned data reuse activities:

- Literature Overview and Physical Use: A comprehensive literature review will be conducted to identify relevant reports, studies, and documents related to sustainable living, AI applications, and the key sectors of manufacturing, energy, health, and transport. Physical use of existing public datasets from previous research projects, plain data sets from other researchers (images, videos, metadata, etc.), governmental reports, and industry studies will be considered to inform and complement the project's initiatives. Possible examples may include but are not limited to:
  - ✓ *Reports on Sustainable Practices in Manufacturing*: provide a description and characterization of sustainable manufacturing practices, schematic representations, diagrams, policy descriptions, and assessments related to the manufacturing sector.
  - ✓ *Health Sector Surveys*: Format: reuse health-related survey data to understand societal perspectives on AI applications in healthcare, contributing to the project's objective of educating society about AI benefits.

In the research background, the consideration of aggregating existing data, especially after anonymization, is a critical aspect. It involves balancing the potential benefits of data aggregation

with ethical, legal, and privacy concerns. The possible existing data will be used after the consideration of the followings: ethical considerations, legal compliances, anonymization techniques used, data transparency, security measurements, and data quality.
Metadata will be enriched to reflect additional processing, analytics, transformations, and interpretations applied to the reused data within the project lifecycle. Detailed documentation of the modifications and enhancements made to the original datasets will be maintained to ensure transparency and traceability.

Repurposed datasets will be documented and made discoverable through data repositories, catalogues, and project wikis. Clear references to the sources and origins of the reused data will be provided to facilitate proper attribution.

Access levels for reused data will be defined based on data sensitivity and permissions. Some datasets may be openly accessible, while others may have restricted access or be subject to embargo, aligning with ethical considerations and data sharing agreements.

2.2 Data collection

SustAInLivWork will collect both existing data from partners and third parties and will create new data within the project. It will collect and produce three categories of data mentioned previously: Project management documents, Educational skills related materials, and Scientific data.

The dataset is collected to facilitate stakeholder engagement, to consult stakeholders and consumers about their views concerning innovations developed, and to help improve and adopt the innovations.

Different category of data will be stored differently:

1. **Project management documents** will undergo a specialized storage and sharing protocol. In contrast to the centralization of scientific data within KTU servers or the open-source dissemination of educational materials, Project management documents will be stored locally at each beneficiary institution and TEAMS platform (for main management documents sharing among the partners). This decentralized storage approach ensures accessibility to pertinent documentation while respecting the autonomy and operational norms of each beneficiary. The sharing of official documents amongst beneficiaries and government institutions will be facilitated through secure and authenticated email channels for communication. This initial mode of exchange provides a reliable means of communication while upholding document security. Each beneficiary institution will maintain its repository of formal documents, contributing to a distributed yet interconnected network of information. The assurance of security remains paramount throughout the exchange process. As the project progresses, considerations may arise prompting the exploration of alternative sharing platforms, which will be introduced if necessary. These platforms will be selected based on their alignment with security standards and the evolving needs of collaborative document sharing within the project.
2. **Educational skills related materials**, distinct from scientific data, will undergo a specific management protocol. Unlike scientific data, these materials will not be housed within the KTU servers but are envisioned to be openly accessible upon completion. The approach entails a systematic preparation and dissemination strategy that ensures the quality, coherence, and appropriateness of the educational content. Regarding the principle" as open as possible as close as necessary", the educational materials will be prepared before being made publicly accessible. This preparation phase involves based on the need an adherence to established protocols or (and), quality standards, or (and), the incorporation of pedagogical best practice. Content will undergo a thorough review to guarantee accuracy and relevance, and appropriate formatting will be applied to enhance accessibility. Following this preparation phase, the

educational materials will be disseminated with due consideration to the optimal timing in alignment with project milestones. Importantly, dissemination will be contingent upon the completion of the preparation phase, ensuring that the materials are robust, accurate, and ready for public consumption. Moreover, the chosen dissemination model emphasizes open-source principles, facilitating broad accessibility and usability. By adopting an open-source approach, the educational materials aim to contribute to the broader educational community, fostering collaboration and knowledge-sharing.

3. **Scientific data**, owing to its nature and sensitivity, will be systematically collected and housed within the servers hosted by KTU. The meticulous handling of this data involves categorization into two distinct types, namely: Ready-to-Use Data (specifically, AI-relevant Platform data) and Raw/Life Data.
    a. *Ready-to-Use Data*, pertaining to AI-relevant Platform data, will undergo a comprehensive preparation process adhering to established protocols. Subsequently, this data will be transferred to a dedicated repository (AI relevant data platform for the AI Research HUB), a deliverable (D2.3) integral to the project. As part of this preparation, the datasets will undergo anonymization, conforming to established quality standards. The repository will not only serve as a storage facility but will also encapsulate research protocols, methodologies, and all requisite documentation, rendering the data ready for utilization in AI applications.
    b. *Raw/Life Data*, characterized by its intricate nature and substantial size, will find storage within KTU servers. The decision to store this data centrally is necessitated by its complexity, rendering local storage within beneficiary institutions unfeasible in certain instances. Raw data, when suitable for archiving within beneficiary institutions, will only be transferred to KTU servers after thorough preparation for repository storage. Throughout this process, data management plans established by local beneficiary institutions will be followed, ensuring the preservation of data integrity, sensitivity levels, and security.

After the project is finished data reuse will be guaranteed over the KTU archive. All data will be stored with back up for 10 years. The archiving of data, especially when dealing with diverse sensitivity levels, requires a nuanced and case-by-case approach to ensure both compliance with ethical standards and the proper utilization of information. In this context, where medical records are involved, considerations extend beyond routine archiving protocols, involving additional approval processes, such as those conducted by Bioethics committees.

2.3 Characteristics of AI-related research data

This category comprises data generated by user interaction in co-creation sessions with the SustAInLivWork project team (PT) as well as in interaction with the existing tools, the digital knowledge platform, the digital learning materials the discussion forums or AI. Research data management refers to the organization, documentation, preservation, discoverability, and reuse of data over its entire lifecycle. It involves the principles, practices, and workflows required for effective stewardship and administration of research data.

Proper research data management allows for data to be more easily discovered, interpreted, replicated, validated, and built upon in future studies. It increases transparency and accountability in research while also saving time and resources compared to managing data haphazardly.

Some of the main components of research data management include data planning, data collection, data processing, providing metadata and documentation, quality assurance, data annotation, data preservation, securing and backing up data, complying with policies and legal requirements, determining future access/restrictions, and preparing data for sharing/publication.

Multiple stakeholders are involved in research data management. Researchers generate and analyse the data, research institutions provide policies and infrastructure, librarians and archivists assist with curation, publishers enact requirements for data availability, and funders often mandate data management plans.

A range of data repositories, metadata schemas, best practice checklists, data visualization tools, support services through libraries, and more infrastructure is being developed to assist researchers and institutions with managing their research data in responsible ways.

The key is enabling data to be FAIR - Findable, Accessible, Interoperable, and Reusable. Overall, research data management promotes knowledge discovery and translation for the scientific community and wider public.

AI can automate the collection of structured and unstructured data from a variety of sources. For example, natural language processing to extract data from literature or object recognition for gathering data from images.

2.4 Data formats

The SustAInLivWork will collect throughout activities qualitative and quantitative data. The project is committed to adopting a principled approach to data formats, prioritizing the utilization of common, widely accepted, and established standards wherever feasible. This commitment is driven by the recognition that widely used data formats enhance interoperability, facilitate data sharing, and contribute to long-term accessibility. By adhering to established standards, the project aims to ensure that its datasets remain compatible with a broad range of tools, platforms, and future research. This approach promotes data integrity and reusability within the project. The table below provides an examples of presentation formats for data.

Table 1. *Examples of data formats*

File format	Explanation/Justification
Microsoft Word (DocX), ODT or TEX (LateX)	Manuscripts will consist of reports generated during the project, all deliverables, publications and internal documents. While PDF will be used for the final version, intermediate version will be used for editabledocuments.
PDF	Fixed document format to be used for alldocuments available for public consumption (e.g. download from the website), dissemination materials (except for video andaudio) and for all deliverables.
XLSX, CSV, ODT, RTV	Format for tables.
ICS,	Data formats that allow users to store and exchange calendaring and scheduling information for events or the cross-country or cross-border meetings.
ZIP, RAR, .jar	Archive formats commonly used to store and large files, plugins, and libraries in binary form.
java, .jls	File format depending on the language chosen for software development.
JPG, PNG, SVG, EPS	File formats for graphics, photographs, and other images.
MP3	Format for audio files.
OGG, FLV, MP4, MOV, AVI	File formats for videos.
DICOM	File format for medical images and related information

3 DATA MANAGEMENT GUIDELINES

Guiding consideration when deciding on openness and public availability of the data will be done on the following basis:

- The available data should be meaningful to the intended users. This in turn will make identification and use easy for them.
- Publicly available deliverables are available via the SustAInLivWork website (https://www.sustainlivwork.eu/). Guidelines of presentation of the materials are provided in D5.1 Communication, Dissemination and Exploitation plan.
- If the GA and the CA permit it and if data is not subject to the GDPR data collected or processed will be made available for access or download on the project website.
- Data with no relevance for persons outside the project consortium may be held back from publication. Each beneficiary will decide on which data is considered to not be of interest beyond the project consortium according to the data categories mentioned previously.
- The project recognizes the paramount importance of ensuring compliance with the GDPR in all aspects of data management, particularly in the context of data intended for publication. As a foundational principle, the project has established a comprehensive institutional process to assess the GDPR compliance of data before its publication. This review process operates at the institutional level, empowering each participating institution to make informed decisions regarding the publication of data under its purview. The decentralization of this decision-making process allows for a nuanced understanding of the specific considerations and nuances relevant to each institution's data. This approach acknowledges the varying nature of data across institutions, considering factors such as sensitivity, consent, and the specific legal and ethical frameworks governing data management.
- Beneficiaries will oversee and assess the adequate implementation of the open data availability.
- Once the internal quality assurance processes have been completed successfully, consortium members are not responsible for maintaining the data or providing support to prospective users. If inadequacies are discovered despite the quality control, authors must correct these.
- Scientific data stored in KTU servers and dedicated repository, a deliverable integral to the project, will only be accessible and open to consortium members. To ensure data security and integrity, full access to the data through an internet interface is restricted solely to verified consortium members who are registered users. They can also view, sync, and share files across different devices operating on Microsoft OS – all under the user's control.
- KTU, who is hosting the repository, as the Project coordinator maintain an overview of the use and the folder structure. Users working with the SustAInLivWork repository benefit from remote collaboration facilities as well as coordinated scheduling.
- Data on the dedicated repository is regularly updated, checked for quality and consistency, and back upped. KTU, who is hosting and operating the platform, and the PT maintain an overview of these processes.
- The PC will widely communicate with the beneficiaries, and if applicable, to the network members updates of the DMP and related processes.

4 FAIR DATA

The following describes how SustAInLivWork plans to make data generated and/or collected by the project FAIR principles. The FAIR principles aim to organize data in such a way as to increase their transparency and standardization, thus making them available long-term. The principles focus both on humans, as well as on machines.

4.1 Making data findable, including provisions for metadata

To uphold the FAIR principles, each Scientific data generated within the project will be assigned a globally unique identifier, ensuring its distinctiveness and traceability across diverse platforms and repositories. Additionally, extensive metadata, including the dataset identifier, will be incorporated to provide comprehensive information about the data, facilitating its discoverability and interpretation. By adhering to these FAIR principles, the project aims to enhance the visibility, accessibility, and usability of its datasets, fostering collaboration and knowledge exchange within the wider research community. In terms of data sharing, the project adopts a case-by-case approach. The PC in collaboration with each beneficiary, will jointly determine the datasets that are deemed suitable for public sharing. This decision-making process recognizes the diverse nature of data, considering factors such as sensitivity, proprietary considerations, and legal or ethical constraints. Due to the early stage of the project, it is not yet possible to define the procedures for dataset generation, as they depend on the typology as well as the number of variables.

For data identification, the partners will use a UUID application to automatically generate and assign a unique ID to each dataset. Public documents and data deemed for long-term use will be associated with a DOI, which guarantees uniqueness as well as automatic data web retrieval.

4.2 Making data accessible

SustAInLivWork will follow the EC approach to open access for data, namely "as open as possible, as closed as necessary", a principle constantly acknowledged by the beneficiaries when considering which data to make accessible.

Scientific data collected from individual participants will be shared in compliance with the GDPR in the repository. Data Owners/Data Providers are responsible for collecting agreements for data transfer. To this end every time the individual participants shall be informed about the reason and implications of the research, they partake in. In the collection of scientific data related to medicine, the project places a paramount emphasis on the sensitivity of the data involved. One pivotal component of framework is the inclusion of additional approval processes, through the scrutiny of Bioethics committees. Participation in the study is contingent upon obtaining informed consent from all patients involved. Moreover, written approvals will be collected from each participant, documenting their consent to contribute their data to the study.

All project data should be prepared before the official deadline as outlined in the list of deliverable plan and project Gantt chart. Contributing partners are expected to make sure that the data is formatted appropriately, anonymized where possible, and ready for open sharing unless decided otherwise.

Regarding the Educational skills related materials, folders and files should be open to all beneficiaries. If specific data or files are subject to special restrictions (e.g. due to privacy reasons) this data should be password protected.

Deliverables are available via the SustAInLivWork website (https://www.sustainlivwork.eu/). The aim is here to maximize the impact of SustAInLivWork. These results will be shared with a wide audience. Selected data will also be shared through publications in scientific journals and presentations at conferences. SustAInLivWork aims to provide as much of its project results open to the repositories (as an example Zenodo OpenAire repository, https://zenodo.org/).

In addition to the procedures described above the open access policy will be complemented by complying with the provisions of the GA and the CA. This includes the following:

- Unless it is against beneficiaries' legitimate interests, project results must be, as soon as possible disseminated to the public using appropriate means.
- Beneficiaries intending to disseminate a result must give appropriate notice in compliance with the CA. They also must observe the required consent from other parties/beneficiaries in line with the same agreement.
- Beneficiaries must ensure open access to all peer-reviewed scientific publications relating to their results.

4.3 Making data interoperable

SustAInLivWork is a capacity development project with a heavy focus on the development of AI for sustainable living and working. Technical interoperability is limited to the digital repository. The SustAInLivWork will develop a taxonomy for its repository. The beneficiaries will look into this taxonomy to understand if it is suitable for the project.

For human interaction within the project, SustAInLivWork will use common data formats and communication protocols that facilitate collaboration and communication. All materials and tools collected or produced by the project will be available in English, and where appropriate, in local languages.

4.4 Increase data re-use

SustAInLivWork is dedicated to fostering a collaboration and knowledge-sharing by licensing the results it produces under Creative Commons licenses. This strategic approach aims to facilitate sharing and reuse of the project's outcomes, aligning with the principles of open science and promoting accessibility to research findings.

However, the project acknowledges the nature of IP associated with each result. Recognizing the diversity in the characteristics and potential applications of different outcomes, the beneficiary responsible for creating the IP will have the prerogative to determine the specific license configuration on an individual basis for each result.

Simultaneously, the decision-making process regarding data sharing agreements will be undertaken collaboratively by the PC and other beneficiaries, addressing the specificity of each case. While the project is committed to open science principles, it recognizes that not all scientific data can feasibly be made openly accessible. Various factors such as data sensitivity, ethical considerations, or legal constraints may influence the decision to restrict access to certain datasets.

This approach aims to strike a balance between the project's commitment to openness and the practical considerations inherent in scientific research. The collaborative decision-making process allows for a contextual evaluation of the circumstances surrounding each intellectual property and dataset, ensuring that the chosen licensing and data sharing agreements align with the principles of responsible and ethical research conduct.

This section will be continuously elaborated throughout the project when more information on the datasets is made available.

5 ALLOCATION OF RESOURCES

All partners are responsible for FAIR data usage and responsible management. A part of the overall WP budget and person months has been dedicated to these activities including the DMP. For the time being, all beneficiaries are responsible for FAIR data management.

Selection criteria for other data to be stored long-term as well as the identification of the meaningful storage period, associated costs and value expected from sharing will be discussed by the whole consortium during Projects regular meetings.
Costs related to open access to research in Horizon Europe are eligible for reimbursement, subject to the specific conditions set out by the GA of the project. Project beneficiaries will be responsible for applying for reimbursement of costs to making data accessible to others beyond the consortium.

6 DATA SECURITY

SustAInLivWork will follow relevant regulations for data protection and data security, particularly the Directive 95/46/EC (EU GDPR).

Access for the Scientific data to the storage and non-public data is possible only (a) for persons nominated by the beneficiaries, (b) persons who registered and validated their data via a two-factor authentication.

Every beneficiary is responsible for the data they produce and store. They are also responsible for ensuring that the data complies with the European Union data protection laws. The overall responsible project coordinator KTU will ensure a backup of critical data. After the completion of the project all the responsibilities concerning data recovery and secure storage will go to the database/facility storing the data.

In addition, the following provisions apply for a Scientific data:

1) Files should not be transferred by e-mail. They should be uploaded on the storage system and the corresponding link sent out to the partners. Erroneous recipients cannot access the system if they do not have a validate user, thus accidental data breaches are minimized.
2) If files must be transferred by e-mail, they should be transferred via secure connections, in an encrypted and password protected form. Passwords will not be exchanged via e-mail but in direct communication.
3) Use of external devices (like USBs, external storage disks, SD-Cards etc.) should be limited whenever possible.
4) Team members are responsible for regular data-backup on their organization's infrastructure or on their personal computers.
5) Data should be labelled in a structure and systematic way.
6) As a general guideline, those researchers directly interacting with the data providing/generating participant should only handle raw data.

7 ETHICS

SustAInLivWork is subject to GDPR regulation, as the project will collect personal data for its research activities. Thus, the project must obtain informed consent, through a standardized form, from each person participating in research and project activities. For scientific data, especially medical data, the complexity and potential impact on individuals necessitate a more thorough and explicit consent process. In such cases, a standardized written consent form will include detailed information about the study, data collection procedures, data storage, potential risks, and the measures in place to safeguard participant confidentiality. However, for activities such as webinars, workshops, or non-intrusive engagements, obtaining consent by "ticking a box" will be used as a more pragmatic and efficient approach. This method referred to as electronic or online consent, is be suitable for scenarios where the level of involvement and potential risks are comparatively lower. Participants will signify their agreement by actively indicating their consent through a checkbox or a similar mechanism. The

PC will design and validate the form in the project beginning and subsequently include it in the template's repository available for beneficiaries.

The target group for research participation are key actors of change, including researchers, business, society, teachers, researchers, students, NGOs, public administrations, funding agencies, and policy makers. SustAInLivWork does not involve any persons who cannot give informed consent.

SustAInLivWork will conduct mostly expert workshops, reflective exercises, etc. These mediums will enable the project activities and results. They will involve participants who are not members of the project consortium but who will provide the necessary input for the results and deliverables. Hence, it is important for SustAInLivWork to observe ethical guidelines for the involvement of research participants.

To this end, SustAInLivWork will follow the principles outlined below when engaging with research participants:

- Informed context: each participant, who is not a member of the project consortium, engaged in research will receive an information page about the project, its objectives, its activities as well as the intended scope of research.
- Voluntary nature: each research participant engages voluntarily with the project and has the right to withdraw anytime from the respective activity.
- Limited scope of participation: data collection from each participant will be done only in the strict bounds of the scope of her/his participation. No other auxiliary data will be collected or stored. The participant has the right and opportunity at any time before the publication of the research results to review her/his data and request deletion of the same.
- Respect: all research participants are afforded the same level of respect and consideration, independent of their gender, religion, nationality, or any other individual characteristic. In addition, project activities will be conducted in such a manner as to ensure that participants are protected from harm and discomfort.

SustAInLivWork beneficiaries will be continuously made aware that as Data Owners/Data Providers they should provide raw data only at the lowest possible resolution necessary for analysis.

8 REFERENCES

1. Project Management Institute. 2017. A Guide to the Project Management Body of Knowledge (PMBOK Guide). 6th ed. Newton Square, PA: Project Management Institute.
2. FAIR Principles (2022), https://www.go-fair.org/fair-principles/
3. Wilkinson et. al (2015), The FAIR Guiding Principles for scientific data management and stewardship, www.nature.com/scientificdata
4. Xaquìn Perez Sindin, The Sage Encyclopedia of Research Methods, https://methods.sagepub.com/reference/the-sage-encyclopedia- of-communication-research-methods/i13206.xml , accessed on November 15, 2022
5. European Commission. "Guidelines on Data Management in Horizon 2020." Version 3.0, 2016. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
6. Kinder-Kurlanda, K., Weller, T., Pfeffer, J., Cooper, M., Hashimzada, Y., Jesson, B., ... and Hausberg, J. P. (2020). "The Tao of Open Science for COVID-19." arXiv preprint arXiv:2004.15038. https://arxiv.org/abs/2004.15038
7. European Commission. "Horizon Europe guidance on data management plans." Draft 1.0, 2022. https://ec.europa.eu/info/sites/default/files/draft-guidance-dmp_he.pdf
8. Whyte, Angus and Tedds, Jonathan. "Making the case for research data management." DCC Briefing Papers. 2011. http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm
9. Corti, Louise, Van den Eynden, Veerle, Bishop, Libby, & Woollard, Matthew. "Managing and sharing research data." UK Data Archive. 2022 edition.
10. Support Your Data: A Research Data Management Guide for Researchers (riojournal.com)
11. Microsoft Word - EvaluationRubric_FlemishDMP (kuleuven.be)
12. Data Management Plans (ffg.at)
13. Funder data management requirements | The University of Edinburgh