

SONATA

situation aware orchestration
of adaptive architecture

Project duration: 1st January 2024 – 31st December 2027

Grant Agreement number: 101137507

WP: 1

Deliverable: 1.2. Data management plan (DMP)

Lead beneficiary: KUL

Submission Date: xxx

Type: Report

Dissemination Level: Sensitive

Due date: 30 June 2024

Revision History:

DATE	VERSION	AUTHOR/CONTRIBUTOR	REVISION BY	COMMENTS
XX.XX.2024	1	Alex Binh Vinh Duc Nguyen		

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

© Copyright 2024 SONATA Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from SONATA Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

Table of Content

List of abbreviations and acronyms.....	- 3 -
SONATA Consortium Partners.....	- 3 -
1. Executive summary.....	- 4 -
1.1. Project summary.....	- 4 -
1.2. Purpose of the deliverable.....	- 4 -
1.3. Intended audience.....	- 4 -
2. Guiding principles.....	- 4 -
2.1. Key concept definitions.....	- 5 -
2.2. Roles and responsibilities.....	- 6 -
3. Data summary.....	- 8 -
3.1. Purpose of data generation.....	- 8 -
3.2. Data categories.....	- 9 -
3.3. Data generation methodology.....	- 9 -
3.4. Data format.....	- 10 -
3.5. Data overview.....	- 10 -
3.5.1. Technical data.....	- 10 -
3.5.2. Research data.....	- 12 -
3.5.3. Research results.....	- 13 -
3.5.4. Administrative data.....	- 14 -
4. FAIR data.....	- 15 -
4.1. Making data findable.....	- 15 -
4.1.1. Metadata.....	- 15 -
4.1.2. Search keywords.....	- 15 -
4.1.3. Naming convention.....	- 16 -
4.1.4. Versioning.....	- 16 -
4.2. Making data accessible.....	- 16 -
4.3. Making data interoperable.....	- 17 -
4.4. Increase data re-use.....	- 17 -
5. Allocation of resources.....	- 17 -
6. Data security.....	- 18 -
6.1. Data security as specified for SharePoint.....	- 18 -
6.2. Data security as specified for Zenodo.....	- 19 -
7. Ethics.....	- 20 -
7.1. Informed consent.....	- 20 -
8. Conclusion.....	- 21 -

List of abbreviations and acronyms

DMP	Data Management Plan
EU	European Union
GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Interoperable, and Reusable
EC	The European Commission
WP	Work Package
IP	Intellectual Property
EEA	European Economic Area
SEN	Sensitive

SONATA Consortium Partners

No	Partner	Short name	Country	Organisation type
1	Katholieke Universiteit Leuven	KUL	Belgium	University
2	Universita Degli Studi di Perugia	UNIPG	Italy	University
3	Inovacijsko-Razvojni Institut Univerze Ljubljani	vIRIUL	Slovenia	Research organisation
4	Universitaetsklinikum Aachen	UKA	Germany	University
5	Technion - Israel Institute of Technology	TECH	Israel	University
6	Masarykova Univerzita	MUNI	Czechia	University
7	Van Berkel En Bos U.N. Studio BV	UNST	Netherlands	Industry
8	Elettrica Valeri SRL	EVALT	Italy	Industry
9	BGrid B V	BGRID	Netherlands	Industry
10	Delta Light	DELTA	Belgium	Industry
11	Rockwool AS	ROCK	Denmark	Industry
12	Green Building Council Italia	GBC Italia	Italy	Organisation
13	Conseil des Architectes D'Europe	ACE	Belgium	Organisation
14	Workplace Innovation Europe CLG	WIE	Ireland	Organisation
15	BLDG System	BLDG	Netherlands	Industry
16	Vetrotech Saint-Gobain International	SAGE	Switzerland	Industry

1. Executive summary

1.1. Project summary

Most workers are dissatisfied with their workplace design, which harms their health, well-being, productivity, and social relations. Although various adaptive workplace technologies aim to manage these risks, their individual and joint impact on worker health and well-being is still unclear. The EU-funded **SONATA project** therefore aims to generate evidence-based recommendations on how architectural adaptation can benefit human health and well-being in various hybrid workplace contexts. The project will measure, quantify, and expand the range of health and wellbeing benefits of architectural adaptation, generate empirical knowledge on how multiple adaptations should be combined to maximise their benefits, and ensure that these benefits are distributed in an equitable way among the workers in a shared workplace.

1.2. Purpose of the deliverable

This deliverable presents the first version of the **SONATA Data Management Plan (DMP)** and describes:

- The guiding principles for data management in the project.
- The legal framework constituted by the General Data Protection Directive (GDPR).
- The data summary, which gives an overview of what data will be gathered and processed by SONATA, i.e. categories, contents, formats, volumes.
- The strategies of how data will be processed in line with the FAIR principles (findable, accessible, interoperable, and reusable).
- The guidelines to assure data security.
- A summary of ethics and privacy strategies regarding SONATA data processing.

The main purpose of the DMP is to ensure good data handling throughout the project, especially providing guidelines on which SONATA data can be shared with the public and how to make them available. The DMP is a living document and will be updated periodically to reflect the actual research data generated during the project, together with updated instructions for how to access open data.

1.3. Intended audience

The dissemination level of D1.2 is ‘*sensitive*’ (SEN) and it is available to the members of the Consortium and the European Commission (EC) services. This document is primarily intended to serve as an internal guideline and reference for all SONATA beneficiaries, especially the governance bodies such as the General Assembly, the Steering Committee, the Industry Steering Board, and the Work Package Leaders.

2. Guiding principles

The guiding principle of the SONATA project is to be as scientifically open as possible, with 87 out of its 121 deliverables being publicly available. Of the 34 that are not public, 24 are administrative plans and reports, and the remaining 10 are technological prototypes.

The handling of SONATA data will conform to the provisions set forth in: **a)** the GDPR (Regulation (EU) 2016/679), **b)** the Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data, and **c)** national laws applicable to the acquisition of valid consent.

As of April 2024, the General Data Protection Regulation (GDPR) is applicable in all Member States in the European Union, as well as in the countries in the European Economic Area (EEA). GDPR updates and modernises existing laws on data protection to strengthen citizens' fundamental rights and guarantee their privacy in the digital age. GDPR regulates the processing by an individual, a company or an organisation of personal data relating to individuals in the EU. GDPR grant individuals a set of rights that must be protected by any actor who processes personal data. **SONATA will ensure that its stakeholders assume the highest level of GDPR individual rights.**

Related documents referenced as the guiding principle for SONATA DMP:

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1713190901660>

[2] Horizon Europe Ethics and data protection: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

[3] Guidelines on data protection impact assessment (DPIA) and determining whether processing is 'likely to result in a high risk' for the purposes of Regulation 2016/679 (last revised and adopted 4 October 2017): https://ec.europa.eu/newsroom/just/document.cfm?doc_id=47711

[4] 2011/61/EU: Commission Decision of 31 January 2011 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data by the State of Israel with regard to automated processing of personal data (notified under document C(2011) 332) Text with EEA relevance: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011D0061>

[5] Report on the first review of the functioning of the adequacy decisions adopted pursuant to Article 25(6) of Directive 95/46/EC: https://commission.europa.eu/document/f62d70a4-39e3-4372-9d49-e59dc0fda3df_en

2.1. Key concept definitions

This section provides the definitions of key concepts in this document.

Personal data categories:

- **Personal data** includes 'any information relating to an identified or identifiable natural person'. An 'identifiable natural person', or '**data subject**', is 'one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' [1].
- **Special categories of personal data** (formerly known as 'sensitive data') are subject to more stringent data-protection safeguards. They include 'personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation' [1].

Data processing concepts:

- **Data processing** includes ‘any operation or set of operations which is performed on personal data or on sets of personal data, **whether or not by automated means**, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction’ [1].
- **Pseudonymisation** entails substituting personally identifiable information (such as an individual’s name) with a unique identifier that is not connected to their real-world identity, using techniques such as coding or hashing. However, if it is possible to re-identify the individual data subjects by reversing the pseudonymisation process, data protection obligations still apply. They cease to apply only when the data are fully and irreversibly anonymised [2].
- **Anonymisation** involves techniques that can be used to convert personal data into fully and irreversibly anonymised data [2].
- **Re-identification** is the process of turning pseudonymised or anonymised data back into personal data by means of data matching or similar techniques.

Roles and responsibilities:

- A **Data Controller** is who ‘alone, or jointly with others, determines the purposes and means of the processing of personal data’ [2].
- A **Data Processor** is who ‘processes personal data on behalf of the controller’ [2].

Although **fully and irreversibly anonymised data are no longer considered personal data** [2], anonymisation processes are challenging, particularly where large datasets containing a wide range of personal data are concerned.

2.2. Roles and responsibilities

Only **six** SONATA partners assume the role of **Data Controllers**, who directly collect personal data from data subjects, including study participants and external stakeholders, including:

- Four partners responsible for conducting intervention evaluation studies, who will collect personal data from study participants:
 - Situational track (ST): **UNIPG** (Italy).
 - Orchestration track (OT): **UKA** (Germany).
 - Socio-cultural track (CT): **IRIUL** (Slovenia).
 - Field track (FT): **KUL** (Belgium).
- Two partners responsible for the recommendation validation workshops, as well as dissemination, exploitation and communication activities, who will collect personal contact details of external stakeholders:
 - Recommendation track (RT): **MUNI** (Czechia).
 - Dissemination track (DT): **EVALT** (Italy).

The **Data Controllers** are responsible to ensure that their data processing is in line with (1) the Horizon Europe GDPR principles, (2) the EU and national data protection laws, as well as (3) the ethical requirements from their respective countries and institutions. They are also responsible to **pseudonymise or anonymise** the data appropriately, and **securely store the raw data as well as any re-identification keys** as outlined in [6. Data Security](#)

As shown in Table 1, all SONATA partners share the responsibility for processing the collected personal data throughout the course of the project as joint **Data Processors**, which is necessary for the project implementation.

Table 1. Overview of roles and responsibilities of SONATA partners (Data Controllers and Data Processors) among its tracks.

SONATA partners	Data subjects					
	Study participants				External stakeholders	
	ST	OT	CT	FT	RT	DT
KUL	Processor	Processor	Processor	Controller	Processor	Processor
UNIPG	Controller		Processor	Processor	Processor	Processor
IRIUL	Processor	Processor	Controller	Processor	Processor	Processor
UKA	Processor	Controller	Processor	Processor	Processor	Processor
TECH	Processor	Processor	Processor	Processor	Processor	Processor
MUNI	Processor	Processor	Processor	Processor	Controller	Processor
UNST					Processor	Processor
EVALT					Processor	Controller
BGRID	Processor	Processor	Processor	Processor	Processor	Processor
DELTA	Processor	Processor	Processor	Processor	Processor	Processor
ROCK		Processor	Processor	Processor	Processor	Processor
SAGE		Processor	Processor	Processor	Processor	Processor
GBC Italia					Processor	Processor
ACE					Processor	Processor
WIE					Processor	Processor
BLDG					Processor	Processor

The **Data Controller** of each track is responsible to set out an Agreement of Data Responsibilities between themselves and the joint **Data Processor**, which specify (1) the data flow; and (2) the data processing activities of respective partners. A summary of this agreement, as well as a single point of contact of the Data Controller, will be provided to the respective data subjects. Together with the **Data Controllers**, the **Data Manager**, who is also the **Project Coordinator** (Andrew Vande Moere, KUL) and the **Project Manager** (Alex Binh Vinh Duc Nguyen, KUL) are responsible in monitoring the generation and execution of these agreements.

SONATA will actively involve the **Data Protection Officers (DPOs)** appointed by its partner institutions, as well as its **External Ethics Advisor** (Petr Jemelka, MUNI), to seek for their advises in data processing, privacy, or protection issues. The contact details of the DPOs at the time of writing this deliverable are:

- KUL : Toon Boon (toon.boon@kuleuven.be) via [KU Leuven GDPR helpdesk](#).
- UNIPG : xxx
- IRIUL : xxx
- UKA : xxx
- TECH : xxx
- MUNI : xxx
- Any other?

For each intervention study, all personal data will only be collected upon receiving **informed consent** from the participants, and any participant providing personal data can at any time withdraw their participation and related data from the project. The privacy, ethics, and procedures for obtaining informed consent are described in detail in **Deliverable 1.8. Ethics management plan (Month 6)**, and a short summary is provided in [7. Ethics](#).

3. Data summary

SONATA does not consider to re-use any existing data; and will only collect or process data after receiving the explicit consent from data subjects.

3.1. Purpose of data generation

Data generation is crucial for SONATA to achieve its three Key Objectives:

KO.1: To realise architectural adaptation as a negotiating link between the health- and wellbeing risks and benefits that exist in the workplace, and their physical and mental health outcomes, which can be positively enforced through the purposeful intervention of multi-layered, situational-aware, orchestrated and equitable workplace design principles.

- To accomplish KO.1, SONATA needs to gather sufficient data to develop and evaluate a **situational-aware digital twin** able to estimate actual workplace situations by means of fusing **automated** (quantitative) data about environmental conditions, human occupancy, the positions of assets and wider-scale activities of workers; and **voluntary** (qualitative) data about individual work activities and preferences. This digital twin then informs an **orchestration model** able to optimise the holistic health and wellbeing effect of concurrent adaptations on different layers to be more beneficial than the combination of each individual layer separately. To ensure equitable individualised adaptation, a **human-building interface (HBI)** will allow workers to provide (anonymous) feedback to communicate individual preferences, change mis-estimated situations or fine-tune adaptations to their liking.

KO.2: To define the principles of an evaluation methodology that is based on objective and subjective short-term health and wellbeing indicators, which can capture the influence of architectural adaptation in the shared workplace to estimate the benefits and risks for longer-term health and wellbeing of individuals.

- To accomplish KO.2, SONATA needs to first **evaluate the most relevant health and wellbeing benefits and risks** for prototypical shared workplaces via a thorough literature review, qualitative observations and interviews with actual workers, and interviews and focus groups with SONATA's key target group representatives. To **measure the health and wellbeing impact of architectural adaptation**, SONATA needs to implement intervention evaluation studies through both short-term controlled experiments and longer-term longitudinal experiments, in which the health and wellbeing impact will be **quantified** through both objective and subjective measures such as environmental, behavioural (e.g., actigraphy) and physiological data collection and processing, and via pre- and post-intervention surveys; and also **qualitatively assessed** through ethnography-informed research methods (e.g. observation, in-depth interviews).

KO.3: To create evidence-based recommendations and guidelines that facilitate the effective application of architectural adaptation for workers, work organisation innovators, architectural designers, OSH experts, building certificate and rating consultants and adaptive technology manufacturers.

- To accomplish KO.3, SONATA needs to first synthesise its results into a unified **recommendation binders** with sufficient evidence (e.g. measurement, data log, literature). These recommendations will be **validated internally** through cost/benefit, sustainability, and barriers to implementation analysis, and participatory workshop with key target group representative; before being **externally disseminated** to a wide audience of at least four different target groups: individuals, organisations, policymakers, and industry professionals;

who will be invited by SONATA partners from their respective networks. Therefore, collecting **contact details** of these external stakeholders is crucial.

3.2. Data categories

Figure 1 illustrates the four main data categories collected and generated by SONATA, including: technical data, research data, research results, and administrative data. The definitions of these data categories are as below:

- **Technical data:** all data related to the implementation and operation of the SONATA architectural adaptation technologies (hardware and software), which is comprised of their source code, schematic design, and their (raw) digital logs.
- **Research data:** all data streams from the SONATA research tracks that benchmark individual health and wellbeing impact of adaptation technologies, and then valorise these knowledges into validated recommendations. Within research data, raw non-anonymised and pseudonymised data are **personal data**.
- **Research results:** all processed data that are presented in publications, deliverables, recommendations, output technologies, and business models. Protected research results will feed into patents and licenses following the IP policy, while non-protected results will be made openly accessible as early as possible to comply with Horizon Europe guidelines for Open Science practices.
- **Administrative data:** all (confidential) data that is generated or shared internally within the SONATA Consortium for administrative and management purposes. Within these data, the contact information of external stakeholders collected from Consortium partners is **personal data**.

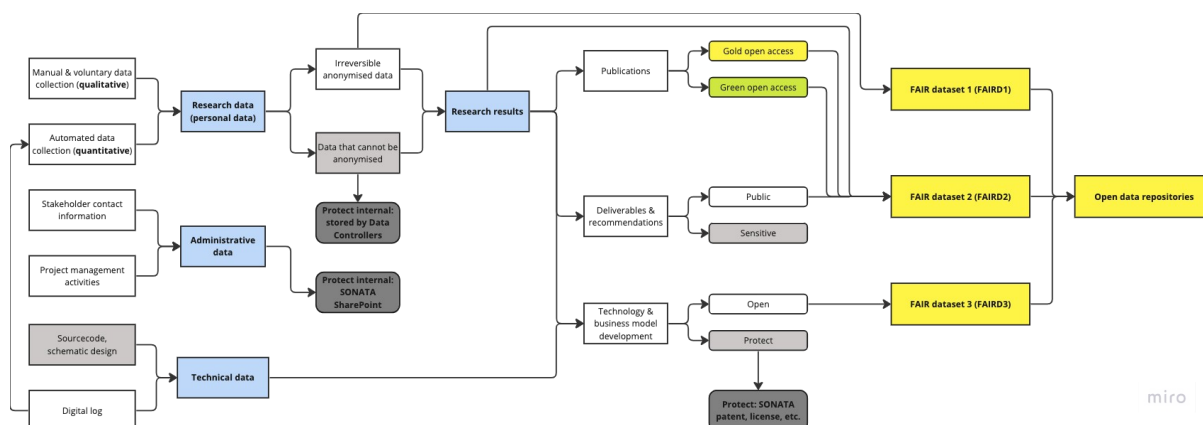


Figure 1. The flow of the four SONATA data categories

3.3. Data generation methodology

Table 2 lists the different data generation approaches that SONATA will employ, resulting from a Consortium-wide survey conducted in Month 5 of the Project implementation.

Table 2. Preliminary overview of SONATA data generation methodology

Generation methodology	Data type	Generation methodology	Data type
Automated digital log	Quantitative	Meeting journaling	Qualitative
Cost / benefit analysis	Qualitative	Mixed-method analysis	Qualitative
Daily communication	Qualitative	Observation	Qualitative
Documentation	Qualitative	Participatory activities	Qualitative
Images & videos of events	Qualitative	Qualitative analysis	Qualitative
Images & videos of studies	Qualitative	Registration to events	Qualitative
Interview recording & transcription	Qualitative	Simulation	Quantitative
Iterative co-authoring	Qualitative	Survey	Quantitative

Iterative design / engineering	Qualitative	Statistical analysis	Quantitative
Literature review	Qualitative	Video recording	Qualitative
Manual digital log	Quantitative	Voluntary feedback	Qualitative

3.4. Data format

SONATA partners are required to collect and generate data in widely accepted data formats, as listed in *Table 3* below.

Table 3. Preliminary overview of SONATA data format

Data format		Description
Document	pdf	Portable format for document and vector image
	txt	Plain text
	docx	Microsoft Words document
	pptx	Microsoft PowerPoint slide show
Database	csv	Comma-Separated Values representation of tabular data
	json	JavaScript Object Notation representation of hierarchical data
	xlsx	Microsoft Excel spreadsheet
Images	jpg	Raster image with lossy compression
	png	Raster image with possible lossless compression
Audio	mp3	Digital audio with lossy compression
	wav	Digital audio with possible lossless compression
Video	mp4	Multimedia container of video/audio
	mov	QuickTime video file
Source code	py	Python programming code
	ino	Arduino programming code
	jv	JavaScript programming code
	html	Markup language defining webpage structure and content
	css	Style sheet language specifying webpage styling
	gh	Grasshopper programming code
Model	3dm	Rhinoceros 3D model
	obj	3D geometry model
	unity3d	Unity 3D package

This list of data formats is not exhaustive, and thus subjected to be changed and/or updated throughout the course of the project. In case any partner employs an unlisted data format, they are required to ensure that such data format (1) is internationally accepted, (2) is accessible to other relevant partners, and (3) follows the FAIR principle (see [4. FAIR data](#)).

3.5. Data overview

An early survey allowed the identification of SONATA data overview, categorised as below. This overview will be updated over the course of the project.

3.5.1. Technical data

Technical data includes all data related to the implementation and operation of the SONATA architectural adaptation technologies (hardware and software), which is comprised of their source code, technical design, and their digital logs. As some technical data is collected via sensors and control interfaces, they may contain personal data. In this case, the GDPR rules apply (see [2. Guiding principles](#)). *Table 4* provides an overview of SONATA technical data at this moment.

Table 4. Preliminary overview of technical data

Adaptation technology	Data controller	Technical data	Data format	Collection methodology	Description	Approximate size (2 days)
-----------------------	-----------------	----------------	-------------	------------------------	-------------	---------------------------

Multi sensor node & access point	BGRID	<i>Environmental sensing</i>	Database	Automated digital log	Including Temperature, Relative humidity, Light intensity, Sound intensity, Passive Infrared, CO2, TVOC	Several KBs
		<i>Occupancy sensing</i>	Database	Automated digital log	Estimation based on data from environmental sensing	Several KBs
Situation-aware digital twin	TECH	<i>Software</i>	Source code	Iterative design / engineering	Digital Twin software	Several MBs
		<i>Sensors and components status</i>	Database	Automated digital log	Data feed from sensors and components	Several KBs
		<i>Occupant input</i>	Database	Manual digital log	Data feed from HBI interface	Several KBs
		<i>Occupant feedback</i>	Database	Voluntary feedback	Data feed from HBI interface	Several KBs
		<i>Visual interface</i>	Model	Iterative design / engineering	Interactive Visualization of Digital Twin data	Several GBs
		<i>Documentation</i>	Document	Documentation	Description of system functionality	Several MBs
Orchestration model	TECH	<i>Software</i>	Source code	Iterative design / engineering	Orchestration software	Several GBs
		<i>Simulation output</i>	Database	Simulation output	Impact of orchestrations on health	Several GBs
		<i>Orchestration model</i>	Model	Iterative design / engineering	AI model identifying optimal adaptations	Several GBs
		<i>Documentation</i>	Document	Documentation	Description of system functionality	Several MBs
HBI interface	KUL	<i>Interface design</i>	Document	Iterative design / engineering	Interface structure, content & styling	Several MBs
		<i>Control software</i>	Source code	Iterative design / engineering	The interface software	Several MBs
		<i>Occupant input</i>	Database	Manual digital log	Occupant control of adaptations	Several KBs
		<i>Occupant feedback</i>	Database	Voluntary feedback	Occupant evaluation of adaptations	Several KBs
Electrochromic glass	SAGE					
Mobile robotic partition	KUL	<i>Technical design</i>	Model	Iterative design / engineering	The overall hardware design	Several MBs
		<i>Robotic software</i>	Source code	Iterative design / engineering	The semi-autonomous software	Several GBs
		<i>Position, orientation</i>	Database	Automated digital log	Real-time location in the 2D map	Several KBs
		<i>Operational status</i>	Database	Automated digital log	Battery, connection, task progress	Several KBs
Kinetic acoustic ceiling	KUL	<i>Technical design</i>	Model	Iterative design / engineering	The overall hardware design	Several MBs
		<i>Robotic software</i>	Source code	Iterative design / engineering	The semi-autonomous software	Several MBs
		<i>Position, orientation</i>	Database	Automated digital log	Real-time estimated position in the 3D space	Several KBs
		<i>Operational status</i>	Database	Automated digital log	Connection, task progress	Several KBs
Dynamic lighting	DELTA					
Height-adjusting desk	KUL	<i>Technical design</i>	Model	Iterative design / engineering	The overall hardware design	Several MBs
		<i>Control software</i>	Source code	Iterative design / engineering	The control software	Several MBs
		<i>Height</i>	Database	Automated digital log	Real-time height estimation	Several KBs
		<i>Occupancy status</i>	Database	Automated digital log	Occupied, vacant, or unknown	Several KBs
		<i>Operational status</i>	Database	Automated digital log	Desk ID, connection	Several KBs
		<i>Occupant ID</i>	Database	Automated digital log	ID from the RFID badge reader	Several KBs
Modular structure	UNST					
Orchestration adaptation control	BGRID	<i>Software</i>	Source code	Iterative design / engineering	Orchestration adaptation control software	Several MBs
		<i>Documentation</i>	Document	Documentation	Description of system functionality	Several MBs

The **Technical Data Controllers**, who are the partners responsible for developing the respective adaptation technologies, will also be responsible for generating, processing, and overseeing the sharing of technical data within the Consortium.

During the implementation of SONATA, the (automated and manual) digital log of technical data will be organised between six databases, as shown in Figure 2. These databases include:

1. **KUL local database.** This database is linked to a central system that monitors and controls the robotic behaviours of three adaptation technologies: mobile robotic partitions, kinetic acoustic ceilings, and height-adjusting desks. This database is [hosted locally by KUL](#), using [Linux MySQL](#), ensuring the data security measures.
2. **BGRID cloud database.** Xxx

3. **UNIPG local database.** Xxx
4. **UKA local database.** Xxx
5. **IRIUL local database.** Xxx
6. **SONATA centralised database:** This centralised database is used to stream and store selected data from the five databases above, considering that these data follow the data privacy and ethics principles. It is [hosted locally by KUL](#), using [Linux MySQL](#), ensuring the data security measures.

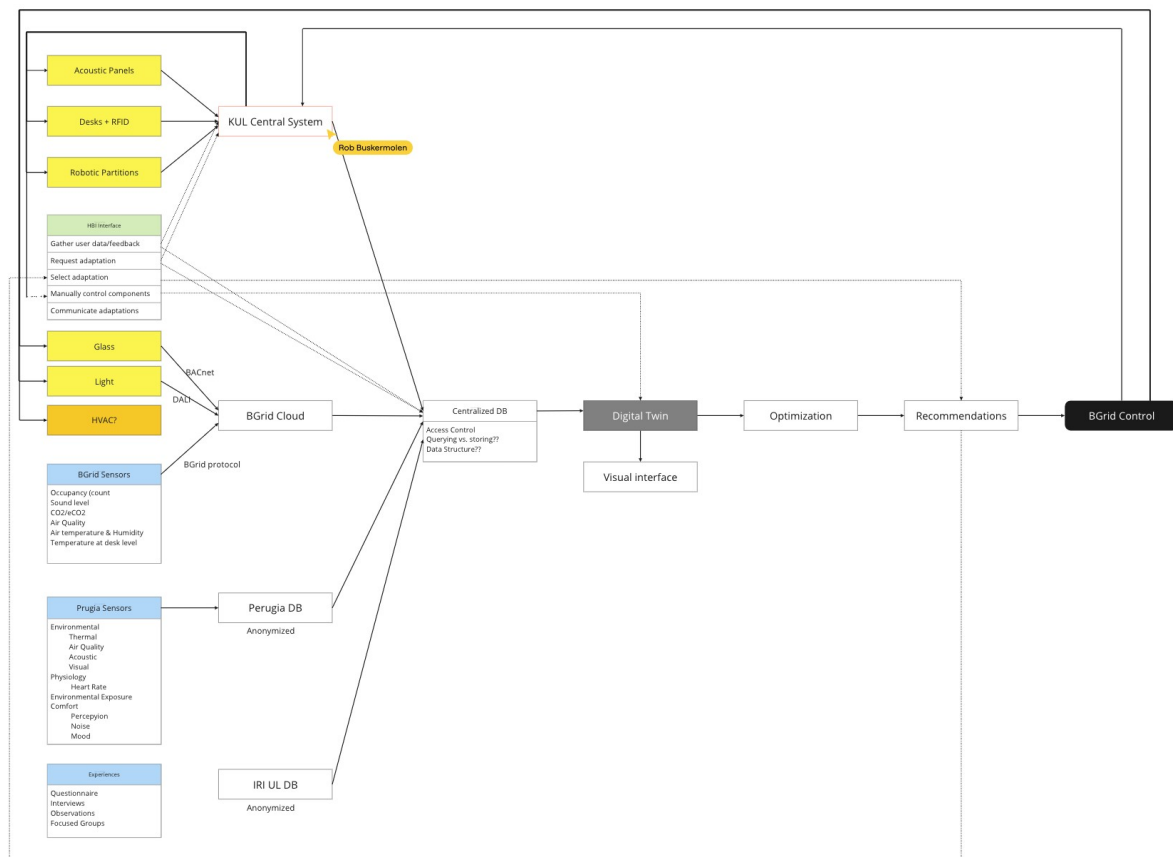


Figure 2. SONATA IT system architecture, showing the data flow of digital log.

3.5.2. Research data

Research data includes all data streams from the SONATA research tracks that benchmark individual health and wellbeing impact of adaptation technologies, and then valorise these knowledges into validated recommendations. As a large portion of research data contains personal data collected from participants, the processing of these data must follow the policy described in [2. Guiding principles](#). Table 5 provides an overview of SONATA research data at this moment.

Table 5. Preliminary overview of research data

Track	Data controller	Research data	Data format	Collection methodology	Description	Approximate size (2 days)
Situational track	UNIPG	<i>Subjective health and well-being</i>	Database	Survey	Weekly questionnaire on perceived sleep quality, stress level, eating and exercise habits in the last week	Several KBs
		<i>Physiological indicators of stress</i>	Database	Automated digital log	This includes measures of heart rate (from FitBit)	Several KBs
		<i>Environmental perception</i>	Database	Survey	Responses gathered from the Cozie app from provided smartwatches (approximately, 4 times/day)	Several KBs
		<i>Environmental satisfaction</i>	Database	Survey	Responses gathered from the Cozie app from provided smartwatches (approximately, 4 times/day)	Several KBs
		<i>Reactions to</i>	Document	Observation	Analysis of pre- and post- interventions focused on	Several KBs

		SONATA solutions			collected subjective and physiological responses	
		Environmental parameters	Database	Automated digital log	Air Temperature, relative humidity, sound pressure level, illuminance at work plan on 5 directions, CO2 concentration, PM1/2.5/10 concentration, radiant asymmetry	Several KBs
		Occupancy	Database	Automated digital log	Level of occupancy at room level from BGRID sensors	Several KBs
		Occupant experience	Document	Interview recording & transcription	Depending on the phase of the study, these will be noted through questionnaire, semi-structured interview, diary logs, focus groups	Several KBs
		Predicted personal comfort status	Model	Automated digital log	This could be an outcome of the study, working on the collected database we aimed at providing individual comfort by correlating physiological, perceptual, and environmental data	Several KBs
Orchestration track	UKA	Subjective health and well-being	Database	Survey	Questionnaires on health and well-being including current symptomatology, affect, life satisfaction	Several GBs
		Physiological indicators of stress	Database	Automated digital log	This includes measures of heart rate, heart rate variability, cortisol levels and potentially EEG data	Several GBs
		Environmental perception	Database	Survey	This includes questions on thermal, visual, acoustic and air quality perception	Several MBs
		Environmental satisfaction	Database	Survey	This includes questions on thermal, visual, acoustic and air quality satisfaction	Several MBs
		Reactions to SONATA solutions	Document	Observation	Behavioural reactions from the participants + potential feedback provided by the participants	Several MBs
		Outcomes from cognitive tasks	Database	Manual digital log	Outcomes from the task activities performed during the experiment	Several GBs
		Medical symptomatology	Document	Observation	Medical examinations by the medical doctor	Several GBs
		Environmental parameters	Database	Automated digital log	Temperature conditions, humidity levels, acoustic levels, lighting levels etc in the lab during the experiment	Several GBs
Socio-cultural track	IRIUL	Subjective health and well-being	Database	Survey	Weekly questionnaire on health and well-being parameters in and outside of the work environment (physical activity, stress level, sleep quality, physiological problems etc.)	
		Physiological indicators of stress	Database	Automated digital log	This includes measures of heart rate, heart rate variability (from FitBit)	
		Environmental perception	Database	Survey	Responses gathered from the Cozie app from provided smartwatches (approximately, 4 times/day)	
		Environmental satisfaction	Database	Survey	Responses gathered from the Cozie app from provided smartwatches (approximately, 4 times/day)	
		Environmental parameters	Database	Automated digital log	Air temperature, relative humidity, sound pressure level, illuminance at work plan on 5 directions, CO2 concentration, PM1/2.5/10 concentration, radiant asymmetry	
		Occupant experience	Document	Interview recording & transcription	Data will be collected through combining qualitative research methods, including semi-structured interviews, diary logs, focus groups	
		Reactions to SONATA solutions	Document	Observation	Participant observation (presence of researcher in the living lab) will be conducted in selected periods of time, to observe and participate in the interactions during interventions and gather behavioural responses and immediate feedback from participants	
		Volunteered personal characteristics	Database	Survey	To focus on different personal traits, relevant to CT (age, gender, dis/ability, personal contexts, such as need for social interaction, specific work requirements etc.), such volunteered data will be collected at the beginning of the study	
Field track	KUL					
Recommendation track	MUNI	Financial specifications	Database	Manual digital log	Technical partners will fill in the form about the financial aspects of the technology	Several KBs
		Workplace specification	Document	Cost / benefit analysis	Users of the CBA tool will fill in the sheet with workplace specifics	Several KBs
		Recommendation binder	Database	Manual digital log	Selected recommendations from the experiments filled into the standardised form	Several KBs
		Literature review database	Database	Manual digital log	Results of the long-term studies relevant for CBA	Several KBs

The **Research Data Controllers**, who are the partners responsible for conducting the SONATA research tracks, will also be responsible for generating, processing, and overseeing the sharing of research data within the Consortium.

3.5.3. Research results

Research results include all processed data that are presented in publications, deliverables, recommendations, output technologies, business models, and FAIR dataset collections.

Specifically, SONATA aims to publish **three collections of FAIR datasets** in the public domain at the end of the project. These datasets include:

1. **FAIRD1. Research data:** the quantitative and qualitative data streams synthesised from the intervention evaluation studies that are already irreversibly anonymised.
2. **FAIRD2. Research results:** the processed and analysed research data that inform any SONATA scientific claim or conclusion, disseminated in the form of scientific publications, numerical data, texts, images, tables, and other relevant formats.
3. **FAIRD3. Technical data:** the generated source code that facilitate the final prototypes of the adaptive technologies - if the IP measures allows.

The publication of the three FAIR datasets will be reported in **D1.4. FAIR dataset publication 1** (KUL, deadline M24), and **D1.5. FAIR dataset publication 2** (KUL, deadline M48).

The selection of specific data to be included in each FAIR dataset will be processed by the following parties:

- The **SONATA Data Manager** (Andrew Vande Moere, KUL), who will propose the overall structure of each FAIR dataset.
- The relevant **Research Data Controllers** (for FAIRD1 and FAIRD2) and **Technical Data Controllers** (for FAIRD3), as detailed in Table 4 and Table 5, who will feed their collected or generated data into the dataset structure at their own initiatives.
- The **Ethics, Gender & Diversity Manager** (Tomáš Ondráček, MUNI), who will perform screenings for privacy and ethics issues in these datasets if requested by concerning partner.
- The **IP Manager** (Marc Hoppermann, BLDG), who will inform the Data Controllers and mediate between relevant partners in case any protected data should not be included in the FAIR datasets.

As shown in *Figure 1*, protected research results will feed into patents and licenses following the SONATA IP policy (**Deliverable 1.10. IP strategy plan - Month 24**), while non-protected results will be made openly accessible as early as possible to comply with Horizon Europe guidelines for Open Science practices, following the **FAIR data strategies** as outlines in section [4. FAIR data](#).

3.5.4. Administrative data

Administrative data includes all (confidential) data that is generated or shared internally within the SONATA Consortium for administration and management, such as meeting minutes, recordings, internal reports, etc., for historical purposes and future follow-up; as well as contact information of external stakeholders collected from Consortium partners.

In the case of **contact information of external stakeholders**, the policy for personal data applies, as detailed in [2. Guiding principles](#). Accordingly, SONATA will inform any external stakeholder about its processing of their contact information, with the following details:

- The overview and objectives of the SONATA project, and why it is relevant for them.
- The methods of how their contact information was obtained, and how it will be processed by SONATA.
- The options to:
 - Correct any inaccurate detail in their contact information.
 - Request the restriction of the processing of their contact detail in any specific case.
 - Request that their contact information to be excluded from SONATA.

These details will be included in a **secured registration form** (Microsoft Form) in the SONATA SharePoint hosted by KUL, allowing individual stakeholders to execute their GDPR rights.

All administrative data is overseen by the management team (KUL), the WP leaders, and task leaders. The data is stored within the SONATA SharePoint platform that requires the authentication of the users. The internal administrative data is thus **confidential**, only accessible for the members of the Consortium.

4. FAIR data

SONATA will manage data in accordance with the FAIR principle (Findable, Accessible, Interoperable and Re-usable data) to maximise access to, and re-use of its research data.

4.1. Making data findable

SONATA will use the [Zenodo](#) repository as the main tool to make our open research data findable. A **SONATA Horizon** community will be established on the Zenodo website, which will host the three FAIR datasets (FAIRD1-3); as well as open technical data. SONATA will then link all Zenodo uploads to the **European Commission Funded Research (OpenAIRE)** community for maximum findability.

4.1.1. Metadata

Metadata associated with each public dataset in Zenodo will by default be as follows:

- Persistent Digital Object Identifiers (DOI)
- Version numbers for version control
- Bibliographic information
- Keywords
- Abstract/description
- Associated project and community
- Associated publications and reports
- Grant information
- Access and licensing info
- Language

SONATA will add the project name and Grant Agreement number.

4.1.2. Search keywords

The Data Controllers of each dataset will be responsible for uploading public datasets that they have generated, as well as assigning specific **search keywords** relevant to these datasets. Dataset-specific keywords must be descriptive to the content of the dataset. For example, a dataset containing information on occupant experience collected in the Field Track should be tagged with corresponding keywords such as: “*occupant experience*”, “*field study*”, “*public workplace*”.

SONATA has also defined a set of general keywords that should apply to all FAIR datasets, scientific publications as well as public deliverables. These are as follow:

- architectural design
- workplace design
- building management
- building automation

- adaptive technology
- interactive architecture
- health and wellbeing
- occupancy comfort

4.1.3. Naming convention

Datasets will be named using the following naming conventions:

SONATA_DS_DataCategoryCode_ResearchTrack_DataController_Description

- “DS” stands for dataset.
- “DataCategoryCode” refers to the list of data categories described in section 3.2:
 - Technical data: TD
 - Research data: RD
 - Research results: RR
 - Administrative data: AD
- “ResearchTrack” refers to the indicator of the relevant research track, if applicable:
 - Situational Track: ST
 - Orchestration Track: OT
 - Socio-Cultural Track: CT
 - Field Track: FT
 - Recommendation Track: RT
 - Dissemination Track: DT
- “DataController” refers to the short name of the partner/Data Controller who is responsible for the dataset.
- “Description” refers to a short description of the content of the dataset.

Example of dataset name:

SONATA_DS_RD_FT_KUL_OccupantExperience
SONATA_TD_FT_KUL_SensorInstallationMap

4.1.4. Versioning

SONATA will employ the Zenodo-supported [DOI versioning](#) for the open datasets. This allows each dataset to be gradually updated after it has been published, during which each updated version will receive a unique DOI, beside a “global DOI” that is assigned to all versions as a single “concept”. For example, DOI versioning of a gradually updated FAIR dataset can look like:

- FAIRD1 version 1 (specific version): *10.7281/zenodo.60943*
- FAIRD1 version 2 (specific version): *10.7281/zenodo.800648*
- FAIRD1 Concept (all versions): *10.7281/zenodo.705645*

4.2. Making data accessible

Subjected to ethical considerations and participant consent, most SONATA data will be **accessible with as few restrictions as possible**. **Protected data** includes personal and sensitive data that is subjected to privacy, commercial, or security reasons. The protocol of how to identify these data is explained in **2. Guiding principles** (for raw personal data), and **3.5.3. Research results** (for processed personal data and data that is required for commercial exploitation).

As mentioned above, all public datasets, scientific publications and deliverables will be uploaded to [Zenodo](#) with persistent identifiers (DOIs) and made openly available, free of charge. SONATA Data Controllers may also publish their open data on other repositories

hosted by their organisations, or relevant to their disciplines, which *must* provide open access as well as persistent identifiers, including such as:

- [KU Leuven RDR – Research Data Repository](#)
- *Any other??? – to be extended by partners*

4.3. Making data interoperable

As SONATA aims to generate *standardised recommendations*, data collection and processing within the project will follow a consistent protocol as outlined in **Deliverable 7.3. Recommendation binder template & taxonomy (Month 6)**. This recommendation binder has been finetuned to reflect the most relevant health & wellbeing standards (e.g. EU-OSHA, DIN, ISO) and/or building rating systems (e.g. WELL, LEED, GBC Italia) that SONATA aims to influence, allowing a standardised Consortium-wide data approach.

As explained in [3.4. Data format](#), SONATA data interoperability is also ensured by using formats that are widely used, preferably non-proprietary formats. Already during the project, data will be stored in a format that is easily accessible to other project partners. Public data will only be shared in commonly accessible formats allowed by the repositories.

4.4. Increase data re-use

In general, the data is collected, generated, processed, and synthesised in accordance with established practices to ensure the data quality. If necessary, guidelines will also be drawn up to ensure consistency of data. The common privacy notice and consent templates, as shown in section [8. Ethics](#) ensure the consistent approach to GDPR compliance.

Specifically, SONATA follows these principles to ensure the re-use of data:

- Data and other outputs that can be made open access should be freely available to all in the repositories/archives and databases.
- Metadata will be included with each dataset as much as possible.
- All public data and other outputs (where applicable) will be licensed under the **Creative Commons license CC-BY 4.0.**, which preserves sufficient intellectual property rights for the authors/creators.
- The dataset will be **documented** as necessary.

Documentation of data ensures the (re)use of data, also during the project but especially when the data is made openly available for all potential re-users in a repository/archive database. The additional documentation describing the data can be presented, for example in *read-me-file* and provided along with the respective data. The format of the documentation, therefore, will follow that of the recommendation binder (**Deliverable 7.3. Recommendation binder template & taxonomy - Month 6**).

5. Allocation of resources

SONATA uses standard tools and free of charge research data repositories. The costs of data management activities are limited to project management costs (**WP1**) and will be covered by allocated resources in the project budget. Long-term preservation of the public data is ensured through Zenodo. Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

The overall responsibility for data management lies with the **Project Coordinator** (Andrew Vande Moere - KUL) and the **Project Manager** (Alex Binh Vinh Duc Nguyen – KUL).

Supporting the coordinator is a data management team consisting of the **Data Controllers** (see [2.2. Roles and Responsibilities](#)), the **Ethics, Gender & Diversity Manager** (Tomáš Ondráček, MUNI), the **IP Manager** (Marc Hoppermann, BLDG), and the **Quality Assurance Manager** (Anna Laura Pisello, UNIPG).

6. Data security

All SONATA Data Controllers and Processors will implement appropriate measures to ensure the protection of personal data, including appropriate research methods and data-processing tools as below:

Do:

- use GDPR-compliant tools to collect, process and store research subjects' personal data.
- take communications security seriously, following the SONATA dedicated protocols as described in this document.
- check the terms and conditions of all of the service providers (software, applications, storage, etc.) to process personal data, in order to identify and mitigate risks to the data subjects.
- encrypt research data and/or the devices on which they are stored, and ensure that keys/passwords are appropriately protected.
- consult the respective DPOs (see [2.2. Roles and responsibilities](#)) or a suitably qualified expert when required for advice on how to achieve a level of data security that is commensurate to the risks to data subjects.

Don't:

- collect data on a personal device such as a smartphone without ensuring that they are properly protected (e.g. consider the implications of automatic back-ups to the cloud, and the device's security features).
- use free services that may use participants' data for their own purposes in lieu of payment, or collect data or communicate with research participants via social media platforms without first assessing the data protection implications.
- use unencrypted email, SMS or insecure 'voice over IP' platforms to communicate with vulnerable participants.
- expose personal data to unauthorised access or use when accessing them remotely (e.g. by using insecure WIFI connections) or travelling to countries where such devices may be inspected or seized.
- assume that research partners, collaborators or service providers have appropriate information security and data protection policies without checking that this is the case.

6.1. Data security as specified for SharePoint

[KUL SharePoint](#) is the online collaboration platform used the SONATA project. A dedicated [SONATA project site](#) has been established on this platform, accessible only by the partner representatives in the consortium. Only anonymous datasets will be uploaded to this SharePoint folder.

The KUL SharePoint site has the following security settings:

- Access level: Restricted to persons (project members only). Further access restrictions on specific folders are enabled.

- Version management as standard provided by Microsoft. In addition, KU Leuven provides a full backup of the data stored on the Sharepoint site in a non-Microsoft data center.
- Encryption: the communication (data transfer) with the SharePoint online cloud storage is done via SSL/TLS. All SSL connections to SharePoint online via the internet are done with 2048-bit keys. Data movements between data centres take place over a private network and are further encrypted. The data is at rest encrypted with BitLocker disk-level encryption combined with per-file encryption of each file. The per-file encryption is particularly strong, given that the encryption of each file is done with a unique encryption key and that each update of the file is done with the same key. These encryption keys are stored in a different location than where the files themselves are. The encryption uses Advanced Encryption Standard (AES) with 256-bit keys and conforms to the American Federal Information Processing Standard (FIPS) 140-2.
- Threat management, security monitoring, and file-/data integrity prevents and/or registers possible manipulation of data.

Documents and elements in the KUL SharePoint sites are stored in Microsoft's cloud solutions, based in Ireland and the Netherlands. There will be no use of data centres outside EU/EEA or in the US.

The **non-anonymised personal datasets** must be stored locally by the Data Controllers and not shared with others, except for **contact information of external stakeholders**, which must be stored in a strict access-controlled SharePoint folder, only accessible by respective **Data Processors** as detailed in *Table 1*.

Retention period. Although personal data should, in principle, be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed, personal data may be stored for longer periods if it is processed solely for archiving purposes in the public interest or for scientific purposes, provided that appropriate technical and organisational measures are taken (arts. 6.1.e and 89 of the GDPR). In this regard, **SONATA partners agreed that the SharePoint data will be stored for a total period of five years after the end of the project.** However, and to the largest extent possible, personal data will be anonymised as soon as possible.

6.2. Data security as specified for Zenodo

The following list describes the security settings for Zenodo:

- **Versions:** Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- **Replicas:** All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period:** Items will be retained for the lifetime of the repository. The host laboratory of Zenodo CERN, has defined a lifetime for the repository of the next 20 years minimum.
- **File preservation:** Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- **Fixity and authenticity:** All data files are stored along with an MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.

- **Succession plans:** In case of closure of the repository, a guarantee has been made from Zenodo to migrate all content to suitable alternative institutional and/or subject based repositories.

7. Ethics

As shown in Table 6, the SONATA project involves at least **3 data processing operations** that may entail higher ethics risks, according to [2]. This section gives a short overview of how these risks will be managed. The detailed measures will be available in **Deliverable 1.8. Ethics management plan (Month 6)**.

Table 6. Assessment of SONATA data processing operations that may entail higher ethics risks and mitigation approaches.

Data processing operations with higher ethics risks		SONATA assessment	SONATA mitigation
<i>Types of personal data</i>	Genetic, biometric or health data	SONATA will collect from study participants both of their subjective health and well-being, as well as more objective health measures (see 3.5.2. Research data).	The collection and processing of this special category of personal data will strictly follow the elaborated protocols outlined in 2. Guiding principles and 6. Data security , as well as any additional principles brought forward by the independent Ethics Advisor and the Ethics Committees of the partner organisations.
<i>Data collection and processing techniques</i>	Using artificial intelligence to analyse personal data	SONATA approach uses AI to determine situations via supervised learning, initially based on synthetic data but then using manually labelled situations reported by participants through surveys and voluntary interactions with the HBI interface. The adaptive orchestrations are selected by using AI-based optimisation methods including reinforcement learning and genetic algorithms (see 3.5.1. Technical data)	SONATA incorporates a transparent AI development by informing participants about their interactions with these AI technologies through a dedicated workshop before the study commencements. This information will be continuously available to participants via an interactive tutorial that is offered on the HBI interfaces (e.g. smartphone app/website, public display), and be linked to all consent forms and informative posters. The direct feedback from the participants , together with the qualitative research results from the SSH partners allows SONATA to constantly identify and rectify unforeseen biases in the input data or the algorithm design. The decision-making of the AI algorithms can always and unequivocally be overridden by human input via the human-in-the-loop approach and the built-in manual controls in each of the adaptive technologies.
<i>Involvement of non-EU countries</i>	Transfer of personal data to non-EU countries	TECH is based in Israel. Its activities require access to the pseudomine personal data of study participants (see 2.2. Roles and responsibilities), yet this data processing does not raise any ethical issues nor relate to any EU Classified Information (EUCI). No studies or physical events are organised in Israel	Israel has received an ' adequacy determination ' from the European Commission, indicating that they have a data protection framework offering a level of protection equivalent to that provided under EU law. In all consent forms, this information will be available, and participants will be asked to give their explicit consent for their personal data to be processed by non-EU partners. Following the protocol as outlines in 2. Guiding principles , Data Controllers are responsible to set out Agreements of Data Responsibilities between themselves and TECH as the joint Data Processor.

7.1. Informed consent

All study participants and external stakeholders will be asked by the respective **Data Controllers** to give their **informed consent** before any of their personal data is processed. The Data Controllers are responsible to keep records documenting the informed consent procedure, including the information sheets and consent forms provided to research participants, and the acquisition of their consent to data processing. These may be requested by data subjects, funding agencies or data protection supervisory authorities.

For consent to data processing to be ‘*informed*’, the data subject must be provided with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language. As a minimum, this should include:

- The identity of the respective Data Controller and, where applicable, the contact details of the DPO (see [2.2. Roles and responsibilities](#)).
- The specific purpose(s) of the processing for which the personal data will be used, including the aim of the SONATA project and the expected results.
- The approach of how the personal data will be processed. Specifically, the informed consent form will explain that personal data will be treated in full confidentiality and will be registered and stored in a secure manner. Collected personal data will always be pseudonymised or anonymised before it is processed. Published data will always be anonymous, and no personally identifiable information will be published in any way.
- The identity of the Data Processors who will have access to the data. The informed consent form will state that data will be handled by a very limited number of authorised personnel within the SONATA Consortium, whose confidentiality will be regulated by legal agreements.
- The subject’s rights as guaranteed by the GDPR and the EU Charter of Fundamental Rights, in particular the right to withdraw consent or access their data, the procedures to follow should they wish to do so, the right to not share their data with partners outside of the EU, and the right to lodge a complaint with a supervisory authority.
- How long the data will be retained before they are destroyed.

8. Conclusion

The approval and release of this deliverable within the consortium constitutes a **formal commitment by partners** to adhere to the data management strategy and the procedures it defines. When the deliverable is approved by the European Commission, this constitutes confirmation that the procedures are considered by the European Commission to be adequate.

As coordinator of the SONATA project, KUL will ensure that any data management issues which may arise during the project will be handled appropriately and in a transparent and fair manner.

The current report will be a living document throughout the project. The DMP will be updated whenever significant changes arise, such as (but not limited to) new data, new innovations, patent filings, changes in the consortium members and others.