



***The Framework Programme for Research & Innovation
Innovation actions (IA)***

Project Title:

Autonomous self powered miniaturized intelligent sensor for environmental sensing and asset tracking in smart IoT environments



AMANDA

Grant Agreement No: 825464

[H2020-ICT-2018-2020] Autonomous self powered miniaturized intelligent sensor for environmental sensing and asset tracking in smart IoT environments

Deliverable

D8.5 Data management plan & ethics v2

Deliverable No.		D8.5	
Workpackage No.	WP8	Workpackage and task type	Project Management
Task No.	T8.3	Task Title	Task 8.3 Data management, Ethics and Standardization
Lead beneficiary		IMEC	
Dissemination level		PU	
Nature of Deliverable		R	
Delivery date		30 June 2020	
Status		Final	
File Name:		AMANDA_D8.5_Data_management_plan_and_ethics_v2-v1.0	
Project start date, duration		02 January 2019, 36 Months	



This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°825464

Authors List

Leading Author (Editor)				
Surname		Initials	Beneficiary Name	Contact email
Kouzinopoulos		CS	CERTH	kouzinopoulos@iti.gr
Co-authors				
#	Surname	Initials	Beneficiary Name	Contact email
1	Bellanger	MB	Lightricity	mathieu.bellanger@lightricity.co.uk
2	Bembnowicz	PI	IMEC-NL	pawel.bembnowicz@imec.nl
3	Kanlis	AK	CERTH	alexkanlis@iti.gr
4	Karanassos	DK	CERTH	dkaranassos@iti.gr
5	Kauer	MK	Lightricity	matthias.kauer@lightricity.co.uk
6	Sideridis	PS	CERTH	sideridis@iti.gr
7	Van de Wiel	RW	IMEC-NL	rik.vandewiel@imec.nl
8	Van Hal	RH	IMEC-NL	roy.vanhal@imec-nl.nl

Reviewers List

List of Reviewers				
#	Surname	Initials	Beneficiary Name	Contact email
1	Brütsch	MB	ZHAW	manuel.bruetsch@zhaw.ch
2	Meli	MM	ZHAW	marcel.meli@zhaw.ch
3	Vujičić	OV	PENTA	oskar.vujicic@penta.hr

Document history			
Version	Date	Status	Modifications made by
V0.3	19/01/2020	Updates to Deliverable D8.3 with new developments on the Data Management Plan and the Data Management Portal	CERTH
V0.6	05/4/2020	Additional updates to the document, including more data files in Section 2.2	CERTH
V0.7	25/04/2020	Added documentation on datasets of the DMP	IMEC
V0.8	01/06/2020	Submitted for internal review	CERTH
V0.9	16/06/2020	Received feedback from reviewers	ZHAW, PENTA
V1.0	30/06/2020	Final version, ready for submission	CERTH

List of definitions & abbreviations

Abbreviation	Definition
ASSC	Autonomous Smart Sensing Card
DG RTD	European Commission's Directorate-General for Research and Innovation
DMP	Data Management Plan
GDPR	General Data Protection Regulation
H2020	Horizon 2020
IIoT	Industrial Internet of Things
IPR	Intellectual Property Rights
LEPPI	Legal, Ethical, Privacy and Policy Issues
MCU	Microcontroller Unit
ORDP	Open Research Data Pilot
PMIC	Power Management Integrated Circuit
PV	Photo Voltaic
RF	Radio Frequency
RIA	Research and Innovation Action

Executive summary

This Deliverable provides a general description related to data management, ethics and standardisation which is applied in the project and focuses on data collected up to M18. The document consists of the following Sections:

- Section 1 provides information about the project scope, the goals of this report and the Legal, Ethical, Privacy and Policy Issues (LEPPI) manager nomination
- Section 2 describes the needs, reasons and methods of AMANDA's data management, together with the implementation of the data management plan during M1-M18 of the project and can be used to track data collected for the project's duration
- Section 3 details ethical aspects of the project

This version includes updated information on data handling and ethics for the AMANDA project during the first 18 months of the project. Further information is provided with reference to the implementation of the initial plan detailed in **Deliverable D8.3 - Data management plan and ethics v1**. Moreover, the Data Management Portal integration is presented and its technical details analysed.

Table of Contents

1. Introduction	8
1.1. Overall technical objectives	8
1.2. Purpose, context and scope of this deliverable	9
1.3. Appointment of a LEPPi manager	10
2. Data management plan	11
2.1. Procedures of data collection	11
2.1.1. Data collection process	11
2.1.2. Data collection description	12
2.1.3. Specification of the required components data collection	13
2.2. Data storage and back-up	13
2.3. Data documentation	14
2.4. Data retention and archiving	23
2.5. Data Management Portal	23
3. Ethical concerns of the AMANDA project	31
4. Conclusions and future work	33

List of Figures

Figure 1 Data management timeline	9
Figure 2 The data roadmap [1]	10
Figure 3 Open access to research data automation.....	24
Figure 4 AMANDA GitHub repository.....	25
Figure 5 GitHub linked to Zenodo.....	25
Figure 6 The DOI Zenodo of an example AMANDA dataset.....	26
Figure 7 Edit field of Zenodo for metadata	27
Figure 8 Basic information fields	28
Figure 9 License and Funding field	29

List of Tables

Table 1 LEPPI nomination	10
Table 2 Data acquisition details.....	13
Table 3 Data storage and data backup information.....	14
Table 4 “DS.CERTH.01.LIS3DH_Accelerometer_10k_resistor” documentation.....	15
Table 5 “DS.CERTH.02.OPT3001_Light_sensor_10k_resistor” documentation.....	16
Table 6 “DS.CERTH.03.SPH_MIC_1k_resistor” documentation	17
Table 7 “DS.IMEC.01.CO2” documentation.....	19
Table 8 “DS.MICRODUL.01.MS1089_SupplyCurrents”	20
Table 9 “DS.MICRODUL.02.MS1089_TemperatureData”	21
Table 10 “DS.MICRODUL.03.MS8892_SupplyCurrents”	22
Table 11 “DS.MICRODUL.04.MS8892_CapacitanceData”	23
Table 12 Suggested retention time with respect to type of data	23
Table 13 Basic information fields	28
Table 14 License and Funding field	29
Table 15 Other Fields.....	30

1. Introduction

1.1. Overall technical objectives

AMANDA is an ambitious project that develops a unique ASSC with the size, feel and look of a credit card. It can be ideal for easy deployments in buildings (smart living environments), as a wearable, for infrastructure monitoring and to mitigate issues related to current and future epidemics. The project covers the triangle of experimentation, development and standardization to optimize the materials behaviour, connectivity, miniaturization, power consumption, security, intelligence, design and cost. AMANDA's partners have the expertise and combination of world-class manufacturing infrastructures and know-how. They are using micro- and nano-technology, new composites, innovative architectures and advanced software. AMANDA's vision is to overcome the existing technological challenges and achieve the development of a user-friendly wearable platform not only for indoor and outdoor environmental sensing, but also for asset- or even people- tracking. A combination of developed and existing off-the-shelf technologies have been selected and are in the process of integration into the ASSC. Innovative PVs (Lightricity PV), PMIC (EPEAS) and batteries (Ilika's solid-state battery), all packed in under a 3mm thickness. The project will introduce technical breakthroughs that will boost further miniaturization, offer increased sensitivity, small footprint and ultra-low power consumption with a maintenance-free lifetime of more than 10 years.

The project execution requires tight cooperation between the partners. This leads to the generation of a significant amount of information, including datasheets, specifications, measurement and reporting data as well as other types of data. The Data Management Plan (DMP) is in place to ensure that all information is categorised and stored in a safe way and can be accessed at any time by authorised users and, in case of public data, by the scientific community. All work done within the AMANDA project aims to incorporate the high standards set by the European Commission. To that end, the consortium put their efforts to make the research data: Findable, Accessible, Interoperable and Reusable (FAIR) to ensure it is soundly managed. This serves as an important factor that leads to knowledge discovery and innovation, while enables easier data and knowledge integration and reuse.

This document is an updated version of **Deliverable D8.3 - Data management plan and ethics v1**, submitted on M6, which set the basis on which the consortium works on the handling of the scientific data produced throughout the duration of the project. In this context, **Deliverable D8.5 – Data management plan and ethics v2** entails the relevant updates, implementations and tools developed by the consortium. The changes introduced in this document include:

- Updates to the outline of the document and modification of the respective Tables and Figures in order to provide the proper relevant information and details as they were explored during M6-M18
- In Section 2.2, three new implementations that contribute AMANDA project's aim towards FAIR data were added
- Section 2.5 was included that demonstrates the way the project's data management portal is integrated and how it handles the data gathered within AMANDA's scope
- The content of Section 4, from Deliverable D8.3, was merged with that of Section 2 to improve the clarity of the document
- Further documentation, Figures and Tables were provided to argue the validity of the project's strategic data handling

An overview of the data management plan's timeline is provided in Figure 1. Between M6 and M18, the DMP was constantly revised based on the new and anticipated developments for the AMANDA ASSC and the result is reflected in this document. During the same period, the Data Management Portal was configured, as detailed in Section 2.5 and different datasets of the project were collected, as discussed in Section 2.3.

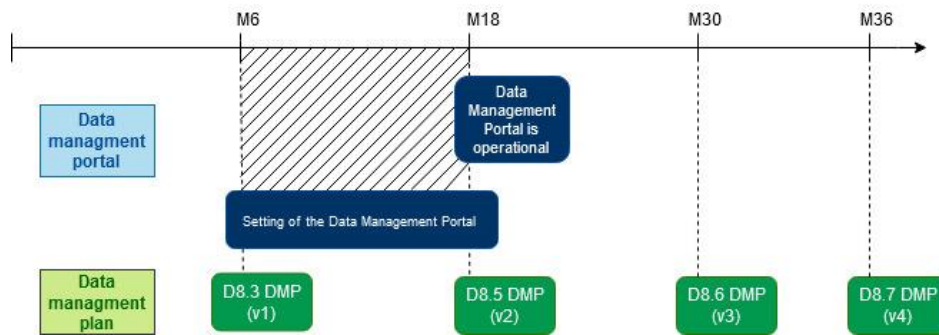


Figure 1 Data management timeline

1.2. Purpose, context and scope of this deliverable

This document relates to data management and ethic plans within the AMANDA project. The DMP is a living document and will evolve during the lifespan of the project. As the project progresses, it will be updated as needed whenever significant changes arise, such as dataset updates or changes in the consortium policies. The reason is that there is a need for handling potentially sensitive data in a discreet way. The data lifecycle of the AMANDA project follows a six-stage process, based on the methodology of [1], as presented in Figure 2 and detailed below:

- **Planning.** The strategizing on how data will be managed, the creation of a data management plan, and being aware of data policies that apply to the work
- **Collecting.** The data acquisition. It refers to two group of techniques: documentation and data organization. Documentation can include metadata, and protocols, as well as documentation files and data dictionaries. Organization strategies for data management include file organization, file naming, how to document conventions, and the use of databases
- **Analysing.** Involves data analysis. It includes understanding the difference between managing raw and analysed data, as well as performing data quality management
- **Publishing.** The data-sharing stage of data's lifecycle
- **Preserving.** Data preservation examines the data to retain and how long to retain it, how to prepare data for the long term, and outsourcing data preservation to data repositories
- **Reusing.** Data reuse is the last stage of data management and the one that completes the data cycle

Subsequent updates to this document are foreseen on M30 (v3) and M36 (v4). The document describes the process applied by the consortium to ensure a quality data management and high ethical standards. It enables clear tracking of data collected, not only during the project execution but also after its conclusion.

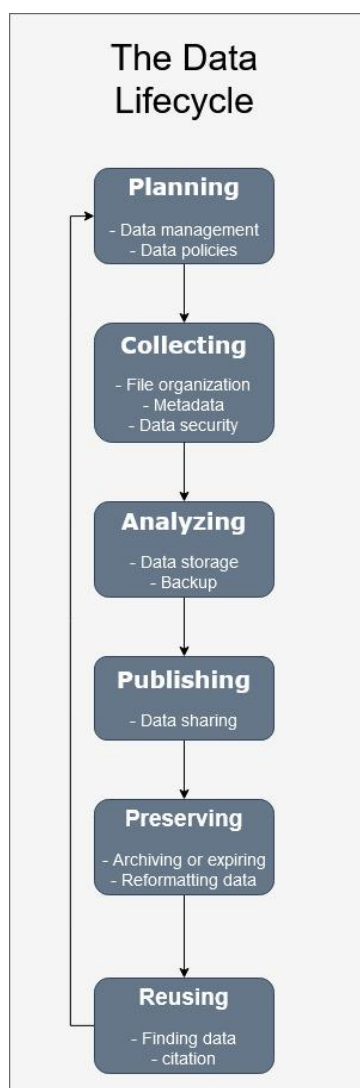


Figure 2 The data roadmap [1]

1.3. Appointment of a LEPPi manager

There is a need to appoint a LEPPi manager. The LEPPi manager is responsible for the coordination of all activities related to legal, ethical, privacy and policy issues that may arise during the development and validation phases of the project. In case of issues related to law, ethics and privacy, the LEPPi manager will cooperate and advise the following decision-making bodies: Plenary Board, Quality Control Board and Ethics Helpdesk. Table 1 shows the assigned LEPPi manager, chosen by the project partners.

Partner short name - company	Name	Email
IMEC	Rik van de Wiel	Rik.vandeWiel@imec.nl

Table 1 LEPPi nomination

Rik van de Wiel is a senior employee at IMEC. He has been working as an R&D manager in the field of connected health solutions for eight years. Rik van de Wiel contributed to many data collection trials, including a number of trials in which medical devices were evaluated.

2. Data management plan

A DMP is a key element of proper data management in a RIA project. It describes the life cycle for the data to be generated, collected and processed. A DMP should include information on:

- The data to be collected, processed and/or generated
- The methodology and standards that will be applied
- The dissemination level of the data
- The way data will be curated and preserved, including after the end of the project's duration

The Commission is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORDP). The ORDP aims to improve and maximize access to and reuse of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. The AMANDA project is part of the ORDP and this Deliverable provides insights regarding the data management, ethics and standardization which are applied within the project and focuses on the data collected up to M18 of the project. The project involves the generation and collection of technical data in different phases, as part of the following Tasks:

- Under **Task T2.2**, off-the-shelf sensors have been evaluated for inclusion to the project. Data produced as part of the evaluation include sensor readings
- Under **Task T2.4**, the output of the data fusion optimisation engine is collected and will be used in the design (**WP3**) and validation phase (**WP5**) of the project
- The entire energy profile of the AMANDA card as it goes through different modes will be captured as part of **Task T2.5**
- Measurements on the interfacing of the power electronics with the system's MCU will be performed under **Task T3.4**
- Under **Tasks T4.2, T4.3 and T4.4**, the output of the cybersecurity, edge intelligence and positioning algorithms respectively will be captured to assist in the integration of the software components with the system's firmware and to enable their optimisation
- The output of **sub-Task T5.1.5** on the testing of the basic firmware routines will be stored to aid in the optional hardware and firmware redesign step under **sub-Task T5.1.6**. The output of the main firmware routines of the system will be recorded under **sub-Task T5.1.7**
- The results of the validation activities of **Tasks T6.2 and T6.3** will be stored for future reference and to assist in the extraction of a set of conclusions that will support the exploitation process of the project and underline key improvements and additions that will optimise the system and increase the targeted spectrum of applications

2.1. Procedures of data collection

This Section summarizes the active efforts of the AMANDA consortium to ensure that technical data generated and collected during the lifetime of the project are findable openly accessible, interoperable and reusable.

2.1.1. Data collection process

The data collection description characterizes, in plain text and for each data collection, the types of research data that will be collected during the study. It also describes how information will be collected and why it is needed. This gives a general overview that can be used to fine tune the data management if required. The following research data types are possible (but not limited to):

- **Observational data.** Captured in real time, typically cannot be reproduced exactly. Examples of observational data include sensor readings, sensory (human) observations, survey results, images
- **Experimental data.** Generated from labs and equipment, can often be reproduced but may be expensive to do so
- **Simulation data.** From models, can typically be reproduced if the input data is known. Examples: output of data fusion algorithms, training datasets of Edge Intelligence algorithms
- **Derived or compiled data.** Theoretical search, data mining or statistical analysis has been done, can be reproduced if the analysis is documented

The data collector can complete the matrix containing the detailed information of the data collection in the study based upon the following check list:

- **Responsible Partner.** Indicates the consortium partner responsible for the data collection
- **Data type.** Describes the type of collected data, including text, arithmetic values, images, models, audio and video files, simulated data
- **Data collection.** Explains the methodology of the data collection: how the data will be collected and from which components
- **Data format.** The format in which the data will be stored, including text and CSV files, compressed and uncompressed images as well as raw binary data
- **Estimated size.** An estimation of the size for each collected data file
- **Software.** Indicates the tool or software needed to create/process/visualize the project-generated data
- **Comments.** Does the data have a specific character in terms of reproducibility, confidentiality and others? What does this mean for the management of the data?

2.1.2. Data collection description

This Section describes the context and type of collected data. Table 2 describes:

- **Dataset name.** It also incorporates the data collector, versioning per collector and details on the way the data was generated. For example, the name “DS.CERTH.01.LIS3DH_Accelerometer_10k_resistor” reflects the fact that the dataset was created by CERTH, it was the first generated dataset by CERTH, it was produced using an LIS3DH accelerometer and the measurement was taken with the integration of a 10k resistor
- **Dataset status.** It includes information regarding whether it’s a new dataset or an updated version of an existing one, the month they are made available and if the dataset is public or confidential

In order to ensure the reproducibility of the results generated, the publicly available data files can be accessed via different software tools, including LibreOffice Calc, Microsoft Excel spreadsheet or MATLAB.

#	Dataset name	Status
1	DS.CERTH.01.LIS3DH_Accelerometer_10k_resistor	New – M15 – Public
2	DS.CERTH.02.OPT3001_Light_sensor_10k_resistor	New – M15 – Public
3	DS.CERTH.03.SPH_MIC_1k_resistor	New – M15 – Public
4	DS.IMEC.01.CO ₂	New – M18 – Confidential
5	DS.MICRODUL.01.MS1089_SupplyCurrents	New – M21 – Confidential

6	DS.MICRODUL.02.MS1089_TemperatureData	New – M21 – Confidential
7	DS.MICRODUL.03.MS8892_SupplyCurrents	New – M21 – Confidential
8	DS.MICRODUL.04.MS8892_CapacitanceData	New – M21 – Confidential

Table 2 Data acquisition details

2.1.3. Specification of the required components data collection

Component data was collected as part of **WP1 - System Specifications, Requirements and Use Cases**. This includes data from components being developed as part of the AMANDA project, such as temperature, touch, CO₂, imaging sensor, solid-state battery, energy harvester, MCU and the PMIC, but also data from State-of-the-Art, off-the-shelf electronic components such as RF chipsets and modules, additional sensors: accelerometer, Volatile Organic Compound, humidity, light as well as other peripherals like timers, displays and memory. Each technological partner within the consortium has contributed to the required data based on its current expertise and on technology scouting e.g. literature survey of patents and datasheets. The collected data was compiled into a CSV document that comprised various Tables of relevant technical specification parameters, especially the electrical and mechanical data, graphs about the power consumption profiles and electronic block diagrams. The purpose of the document is to share the same level of information between all partners, regardless of the individual level of expertise, in order to mutually understand the current status of each respective technology and put these into perspective with the intermediate and final project specification targets. As such, it can be considered as an internal technical project roadmap. Finally, it is also a comprehensive comparison tool for assessing the various sensors, RF and loads that will be integrated into the ASSC. This reference document contains sensible and confidential information that is therefore only accessible by the AMANDA project partners. Any relevant non-confidential information can be subsequently included into the public deliverable documents, for example D1.2 or D1.3. The CSV document will be updated throughout the project and the latest version is regularly circulated to all partners.

2.2. Data storage and back-up

It is the responsibility of the partners to ensure that the data is regularly backed-up and stored securely for the lifetime of the project and beyond. These services should be implemented both for confidential as well as for publicly available data. The following types are distinguished:

- Hosting service (GitHub) with web based graphical interface that provides access control and several collaboration features, such as wikis and basic task management tools, to make public data easily accessed
- Hosting service (GitLab) with web based graphical interface that provides access control and several collaboration features, such as wikis and basic task management tools, for internal co-operation
- A general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN integration (Zenodo)

Backup indicates the location of the backups and the responsible partner. The data in this study will be stored and backed up as described in Table 3.

Ref. nr.	Responsible Partner	Data type	Storage medium and location	Backup location and backup frequency
1	CERTH	Confidential datasets	GitLab	Automated service backup

2	CERTH	Public datasets	Zenodo	Automated service backup
---	-------	-----------------	--------	--------------------------

Table 3 Data storage and data backup information

2.3. Data documentation

The data processed is documented and labelled for immediate usage and future reference. The labelling consists of two parts:

- File naming. Files have naming conventions for each data type. There are many conventions for file naming. The well documented practical guidance from Purdue University [2] is followed. Naming convention is very helpful in case of manual and automatic search
- Metadata. Files have metadata that describes the data stored in the file. Metadata have a description what the data contains and what each value represents. The reason to use metadata is, that it can be found easily when looking for information

Wherever possible, existing community standards should be identified and reused. An example of commonly used generic metadata can be found at Dublin Core Metadata Initiative [3]. The data processed in the study have been documented according to the standards with the data type as described in Table 4.

DS.CERTH.01.LIS3DH_Accelerometer_10k_resistor	
Data identification	
Source	LIS3DH accelerometer sensor
Description	This dataset was gathered in order to establish the baseline on acceleration metrics. It consists of the acceleration forces on three different axis (x,y,z) together with the respective timestamps of each measurement. These measurements were collected during the evaluation of the sensor for Deliverable D2.2
Partners activities and responsibilities	
Partner owner of the device	The device is owned by CERTH, where the data collection was performed
Partner in charge of data collection	CERTH
Partner in charge of data analysis	CERTH
Partner in charge of data storage	CERTH
WPs and tasks	The data was collected within the activities of WP2 and more specifically within the activities of Task T2.2
Standards	
Info about metadata (production and storage dates, place) and documentation?	<p>The dataset is accompanied with a detailed documentation of its contents. Indicative metadata include:</p> <ul style="list-style-type: none"> • Upload type • A Digital Object Identifier

	<ul style="list-style-type: none"> • Publication date • Access rights • References • Contributors
Standards, format, estimated volume of data	CSV data format. This dataset's size is 16kB
Data sharing and exploitation	
Dissemination level	Public
Data access policy	The full dataset is openly available through the AMANDA Data Management Portal
Data sharing, re-use and distribution (How?)	The full dataset is openly available through the AMANDA Data Management Portal
Data exploitation (purpose/use of the data analysis)	The collected data serves as the baseline for the orientation and activity detection of the ASSC. Moreover, this dataset could be useful in the benchmarking of different implementations of any 3-axis MEMS accelerometer to ensure their optimal performance
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): Where? For how long?	The data will be available through the AMANDA Data Management Portal for the duration of the project's lifetime

Table 4 "DS.CERTH.01.LIS3DH_Accelerometer_10k_resistor" documentation

DS.CERTH.02.OPT3001_Light_sensor_10k_resistor	
Data identification	
Source	OPT3001 light sensor
Description	This dataset was gathered in order to establish the baseline on the brightness of the ambient light over time. It consists of the intensity of light as visible by the human eye, in lux, together with the respective timestamps of each measurement. These measurements were collected during the evaluation of the sensor for Deliverable D2.2
Partners activities and responsibilities	
Partner owner of the device	The device is owned by CERTH, where the data collection was performed
Partner in charge of data collection	CERTH
Partner in charge of data analysis	CERTH
Partner in charge of data storage	CERTH

WPs and tasks	The data was collected within the activities of WP2 and more specifically within the activities of T2.2
Standards	
Info about metadata (production and storage dates, place) and documentation?	<p>The dataset is accompanied with a detailed documentation of its contents. Indicative metadata include:</p> <ul style="list-style-type: none"> • Upload type • A Digital Object Identifier • Publication date • Access rights • References • Contributors
Standards, format, estimated volume of data	CSV data format. This dataset's size is 15kB.
Data sharing and exploitation	
Dissemination level	Public
Data access policy	The full dataset is openly available through the AMANDA Data Management Portal
Data sharing, re-use and distribution	The full dataset is openly available through the AMANDA Data Management Portal
Data exploitation (purpose/use of the data analysis)	The collected data serves as the baseline for the brightness of the ambient light over time of the ASSC. Moreover, this dataset could be useful in the benchmarking of different implementations of any ambient light sensor to ensure their optimal performance
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): Where? For how long?	The data will be available through the AMANDA Data Management Portal for the duration of the project's lifetime

Table 5 "DS.CERTH.02.OPT3001_Light_sensor_10k_resistor" documentation

DS.CERTH.03.SPH_MIC_1k_resistor	
Data identification	
Source	SPH0645LM4H microphone sensor
Description	This dataset was gathered in order to establish the baseline on the spectrum frequency over time. It consists of the soundwave amplitude together with the respective timestamps of each measurement. These measurements were collected during the

	evaluation of the sensor for Deliverable D2.2.
Partners activities and responsibilities	
Partner owner of the device	The device is owned by CERTH, where the data collection was performed
Partner in charge of data collection	CERTH
Partner in charge of data analysis	CERTH
Partner in charge of data storage	CERTH
WPs and tasks	The data was be collected within the activities of WP2 and more specifically within the activities of T2.2
Standards	
Info about metadata (production and storage dates, place) and documentation?	<p>The dataset is accompanied with a detailed documentation of its contents. Indicative metadata include:</p> <ul style="list-style-type: none"> • Upload type • A Digital Object Identifier • Publication date • Access rights • References • Contributors
Standards, format, estimated volume of data	CSV data format. This dataset's size is 15kB
Data sharing and exploitation	
Dissemination level	Public
Data access policy	The full dataset is openly available through the AMANDA Data Management Portal
Data sharing, re-use and distribution	The full dataset is openly available through the AMANDA Data Management Portal
Data exploitation (purpose/use of the data analysis)	The collected data serves as the baseline for the spectrum frequency over time of the ASSC. Moreover, this dataset could be useful in the benchmarking of different implementations of any I ² S MEMS microphone to ensure their optimal performance
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): Where? For how long?	The data will be available through the AMANDA Data Management Portal for the duration of the project's lifetime

Table 6 "DS.CERTH.03.SPH_MIC_1k_resistor" documentation

DS.IMEC.01.CO2

Data identification	
Source	The CO ₂ transducer is prepared from commercially available ITO on glass electrodes, covered with an ionic liquid based electrolyte layer
Description	Using the IMEC gas setup, a series of CO ₂ steps of ~2000 ppm were switched on and off in a 5 minute cycle. At the 1h07m mark, the diluting gas for the setup was changed from pure nitrogen to compressed dry air, while at ~2h this was switched back to nitrogen
Partners activities and responsibilities	
Partner owner of the device	The device is owned by IMEC, where the data collection was performed
Partner in charge of data collection	IMEC
Partner in charge of data analysis	IMEC
Partner in charge of data storage	IMEC/CERTH
WPs and tasks	The data were collected within the activities of WP2 and more specifically within the activities of T2.2
Standards	
Info about metadata (production and storage dates, place) and documentation?	In the file detail, a verbal description of the experiment is provided
Standards, format, estimated volume of data	The data format is xlsx, a MS Excel file. The estimated volume of data is approximately 140kB
Data sharing and exploitation	
Dissemination level	Confidential
Data access policy	The data is research measurement. It should remain confidential
Data sharing, re-use and distribution	The created dataset is shared through the Data Management Portal only for members of the Consortium and the European Commission
Data exploitation (purpose/use of the data analysis)	The results show a fast response of both the real (Re) and imaginary (Im) parts of the measured impedance (Z) to changes in the CO ₂ concentration in the absence or presence of oxygen. Previous sensors based on Pt on Si showed (reversible) deactivation when exposed to oxygen
Embargo periods (if any)	End of the project
Archiving and preservation (including storage and backup)	

Data storage (including backup): Where? For how long?	All information belongs to the AMANDA consortium. All data will respect the partner policies
--	--

Table 7 “DS.IMEC.01.CO2” documentation

DS.MICRODUL.01.MS1089_SupplyCurrents	
Data identification	
Source	Microdul MS1089 solid-state temperature sensor
Description	Supply current measurement in prototype package, in a lab environment. The supply current of the temperature sensor is measured during the start-up phase, the idle phase and the measurement phase
Partners activities and responsibilities	
Partner owner of the device	The device is owned by Microdul, where the data collection was performed
Partner in charge of data collection	Microdul
Partner in charge of data analysis	Microdul
Partner in charge of data storage	Microdul
WPs and tasks	The data was collected within the activities of WP2 and more specifically within the activities of task T2.1
Standards	
Info about metadata (production and storage dates, place) and documentation?	The measurement setup and the conditions (voltage, timing, etc.) are described in the data file(s) itself
Standards, format, estimated volume of data	The data format is in CSV. The estimated data size is a few MB
Data sharing and exploitation	
Dissemination level	Confidential
Data access policy	The data is research measurement. It should remain Confidential
Data sharing, re-use and distribution	The created dataset is available only for members of the Consortium. The data can be used to derive simulation models for the energy simulation tool developed in Task T1.4
Data exploitation (purpose/use of the data analysis)	The data is used for evaluation and comparison with the specified power consumption figures, and for modelling the power consumption in the AMANDA ASSC
Embargo periods (if any)	None

Archiving and preservation (including storage and backup)	
Data storage (including backup): Where? For how long?	The data will be stored beyond the lifetime of the project

Table 8 “DS.MICRODUL.01.MS1089_SupplyCurrents”

DS.MICRODUL.02.MS1089_TemperatureData	
Data identification	
Source	Microdul MS1089 solid-state temperature sensor
Description	Temperature data obtained from the sensor in prototype package, in a lab environment. The temperature data of the sensor in a calibrated state is obtained over the full temperature range
Partners activities and responsibilities	
Partner owner of the device	The device is owned by Microdul, where the data collection was performed
Partner in charge of data collection	Microdul
Partner in charge of data analysis	Microdul
Partner in charge of data storage	Microdul
WPs and tasks	The data are going to be collected within the activities of WP2 and more specifically within the activities of task T2.1
Standards	
Info about metadata (production and storage dates, place) and documentation?	The measurement setup and the conditions (voltage, temperature, etc.) are described in the data file(s) itself
Standards, format, estimated volume of data	The data format is in CSV. The estimated data size maximum a few MB
Data sharing and exploitation	
Dissemination level	Confidential
Data access policy	The data is research measurement. It should remain Confidential
Data sharing, re-use and distribution	The created dataset is available only for members of the Consortium
Data exploitation (purpose/use of the data analysis)	The data is used for evaluation, and for comparison with the specified accuracy figures
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	

Data storage (including backup): Where? For how long?	The data will be stored beyond the lifetime of the project
---	--

Table 9 “DS.MICRODUL.02.MS1089_TemperatureData”

DS.MICRODUL.03.MS8892_SupplyCurrents	
Data identification	
Source	Microdul MS8892 ultra-low power capacitive sensor and touch switch for human body detection and system wake-up with absolute & relative switching threshold
Description	Supply current measurement in prototype package, in a lab environment. The supply current of the temperature sensor is measured during the start-up phase, the idle phase and the measurement phase, both with external and internal clock sources
Partners activities and responsibilities	
Partner owner of the device	The device is owned by Microdul, where the data collection was performed
Partner in charge of data collection	Microdul
Partner in charge of data analysis	Microdul
Partner in charge of data storage	Microdul
WPs and tasks	The data was collected within the activities of WP2 and more specifically within the activities of task T2.1
Standards	
Info about metadata (production and storage dates, place) and documentation?	The measurement setup and the conditions (voltage, timing, etc.) are described in the data file(s) itself
Standards, format, estimated volume of data	The data format is in CSV. The estimated data size is a few MB
Data sharing and exploitation	
Dissemination level	Confidential
Data access policy	The data is research measurement. It should remain Confidential
Data sharing, re-use and distribution	The created dataset is available only for members of the Consortium. The data can be used to derive simulation models for the energy simulation tool developed in Task T1.4.
Data exploitation (purpose/use of the data analysis)	The data is used for evaluation and comparison with the specified power consumption

	figures, and for modelling the power consumption in the AMANDA ASSC
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): Where? For how long?	The data will be stored beyond the lifetime of the project

Table 10 “DS.MICRODUL.03.MS8892_SupplyCurrents”

DS.MICRODUL.04.MS8892_CapacitanceData	
Data identification	
Source	Microdul MS8892 ultra-low power capacitive sensor and touch switch for human body detection and system wake-up with absolute & relative switching threshold
Description	Capacitance measurement in prototype package, in a lab environment. The capacitance will be simulated by a programmable voltage source. The linearity of the measurement over the capacitance range will be evaluated
Partners activities and responsibilities	
Partner owner of the device	The device is owned by Microdul, where the data collection was performed
Partner in charge of data collection	Microdul
Partner in charge of data analysis	Microdul
Partner in charge of data storage	Microdul
WPs and tasks	The data were collected within the activities of WP2 and more specifically within the activities of Task T2.1
Standards	
Info about metadata (production and storage dates, place) and documentation?	The measurement setup and the conditions (voltage, temperature, etc.) are described in the data file(s) itself
Standards, format, estimated volume of data	The data format is in CSV. The estimated data size maximum a few MB
Data sharing and exploitation	
Dissemination level	Confidential
Data access policy	The data is research measurement. It should remain Confidential.
Data sharing, re-use and distribution	The created dataset is available only for members of the Consortium

Data exploitation (purpose/use of the data analysis)	The data is used for evaluation, and for comparison with the specified accuracy figures
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): Where? For how long?	The data will be stored beyond the lifetime of the project

Table 11 “DS.MICRODUL.04.MS8892_CapacitanceData”

2.4. Data retention and archiving

This Section describes how long the data will be stored for this study, what data can be archived and what safeguards are setup for the data archiving. Examples of safeguards are limited access, anonymisation, scrambling and deleting parts of data.

The data processed in this study should have the defined retention period for each data type. The default retention period is set to the end of the AMANDA project activities. However, the retention time should be adjusted to significance of collected data. Table 12 shows the suggested retention time with regards different type of data. Nevertheless, the experiment designer should have decisive voice about retention period. Moreover, the files can be archived or deleted directly after processing.

Nr.	Data type	Suggested retention period
1	Non-anonymised raw measurements	Delete subsequent to project end
2	Anonymised raw measurements	5 years or more
3	Project reports	5 years or more

Table 12 Suggested retention time with respect to type of data

2.5. Data Management Portal

In AMANDA, the focus is on the development and operation monitoring of environment, air quality and asset tracking. Two different types of data are collected within the AMANDA project:

- Scientific/technology data from the development side
- Private information, gathered by the consortium, that derives from human participation

The project involves carrying out the data collection and a set of validation tests to assess the technology and effectiveness of the proposed framework in real life conditions. The various models designed in the AMANDA project will be anonymised and will be provided to the public as open models using the Open Model Initiative (OMI). They will be available to interested parties that will provide adaptations and contribute to the evolution of AMANDA models after the end of the project. Hence, it becomes apparent that the majority of research data generated by the project is and will continue to be open. In this context, a detailed data management plan was delivered in M6, as reported in **Deliverable D8.3 - Data management plan & ethics v1** and was subsequently updated in **Deliverable D8.2 - Project progress report** on M10 and in **Deliverable D8.5 - Data management plan & ethics v2** on M18. The plan fully describes the procedures for ensuring that the data management process complies with national and EU legislation. The consortium’s approach is in full compliance with the EU legislative and regulatory framework for data protection based on the uniform approach of the EC Directive

95/46/EC3, and the national legislative and regulatory framework for data protection of each project member country.

In that spirit, an automated way to make the relevant data publicly available was developed. The first component is a dedicated “open access to research data” repository on a self-hosted Gitlab solution. As a next step, this repository is exclusively mirrored to an open-access depository on the Github platform (<https://github.com/amanda-project/ORDP-repository>). The last component of this automation is the Zenodo¹ platform. The aforementioned Github account is integrated with a Zenodo account that makes all the data publicly available and provides a DOI for each item in order for other interested parties to cite them easily. This mechanism is illustrated in Figure 3 below.

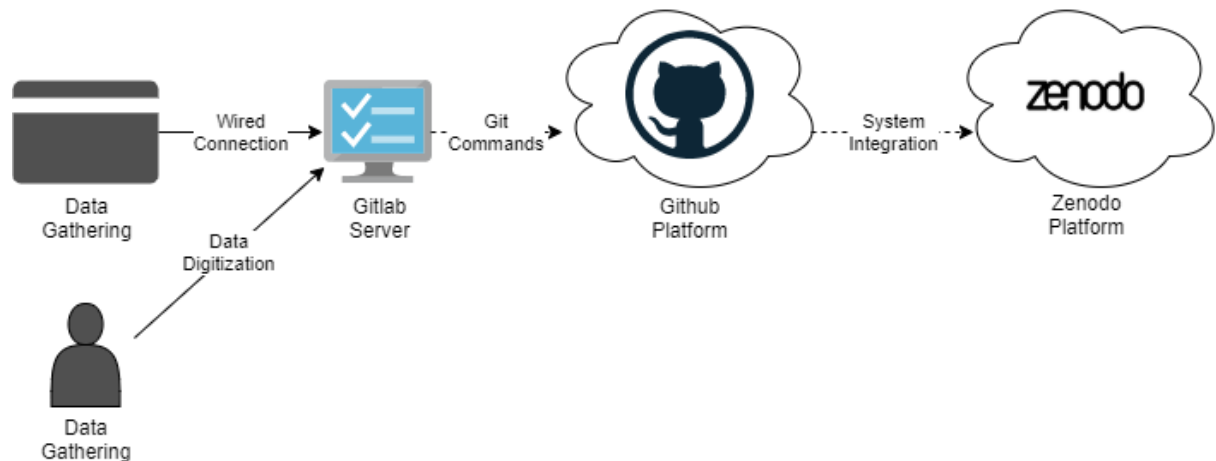


Figure 3 Open access to research data automation

Specifically, there are certain actions that need to be followed in order to enable AMANDA’s repository on GitHub to be citable and have a DOI assigned via the general-purpose open-access repository Zenodo. The first step in providing a DOI, to datasets and other submitted data which lacks one, is to select the repository of GitHub that will be integrated with Zenodo. An overview of AMANDA’s current GitHub repository is given in Figure 4.

¹ Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. It allows researchers to deposit data sets, research software, reports, and any other research related digital artifacts. For each submission, a persistent digital object identifier (DOI) is minted, which makes the stored items easily citable.

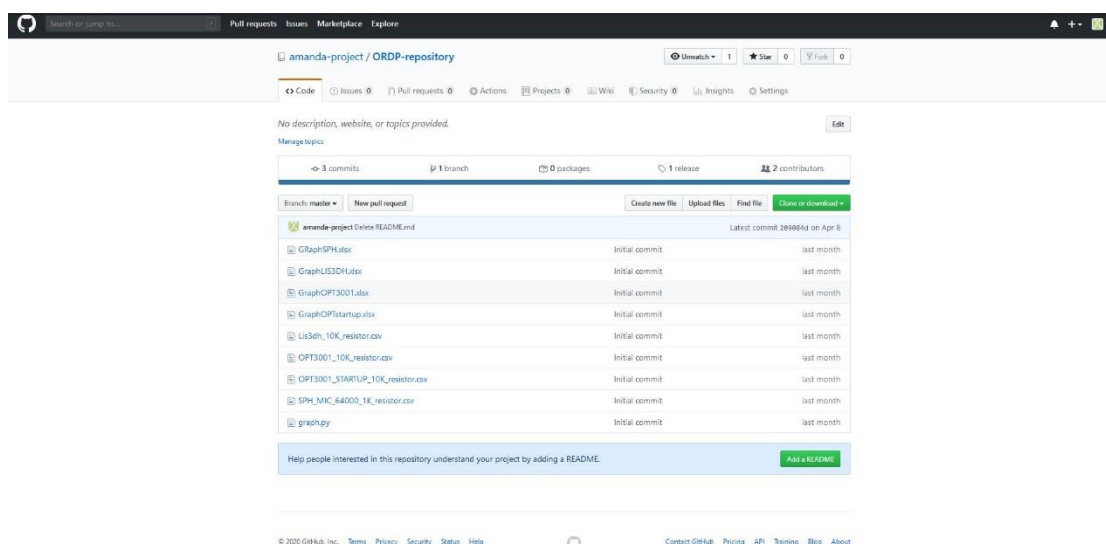


Figure 4 AMANDA GitHub repository

As the project progresses, the repository will be updated. Moreover, the GitHub account needs to be linked to Zenodo.

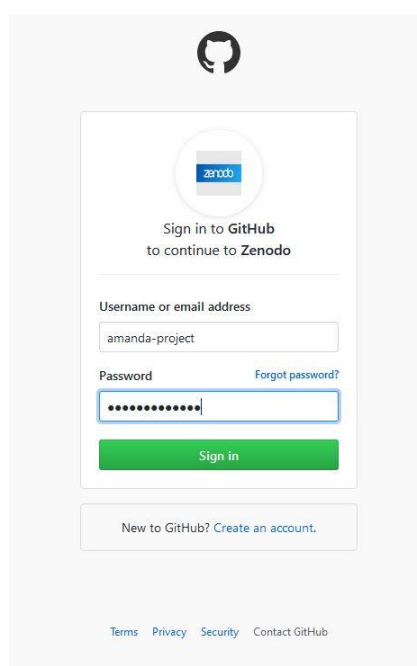


Figure 5 GitHub linked to Zenodo

Furthermore, a new release needs to be created within the GitHub project and Zenodo takes care of the automation that deposits the files and assigns a DOI to the relative data. As a result, AMANDA project's data sets, research software, reports, and any other research related digital artefacts will be stored and have a persistent (DOI) which makes them easily citable. The Zenodo DOI of an example AMANDA dataset can be seen in Figure 6 below.

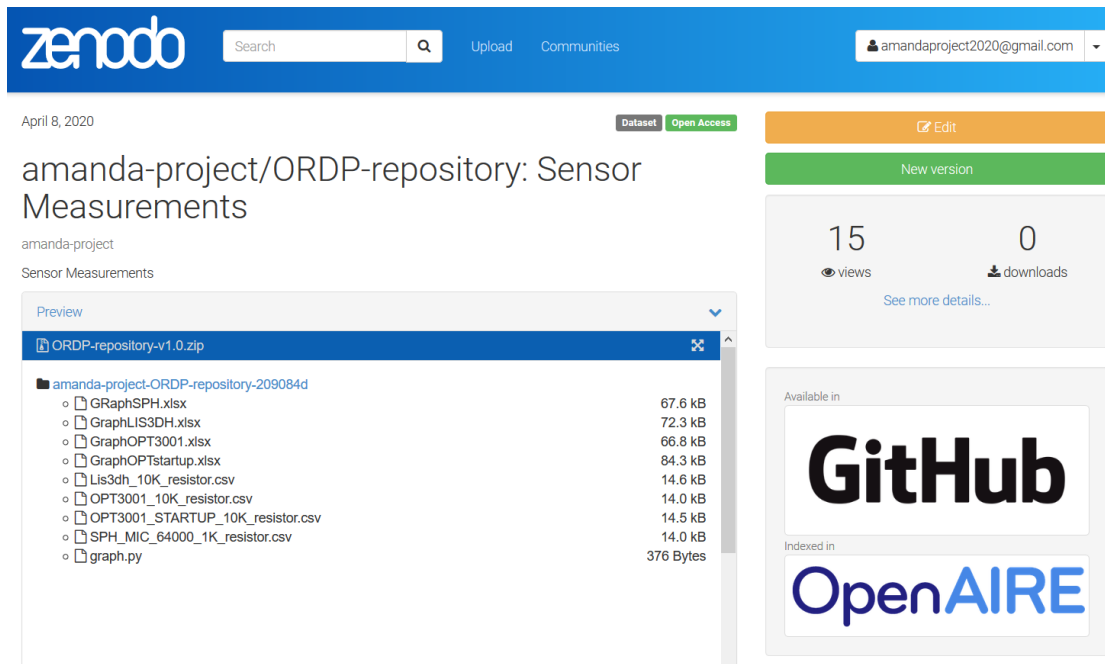


Figure 6 The DOI Zenodo of an example AMANDA dataset

The opportunity offered by this automation for AMANDA is to:

- Manage the datasets which will be collected within the project's actions
- Be flexible in terms of the parts of datasets that are made publicly available
- Facilitate the management of the data produced for the purposes of the AMANDA Project
- Provide a web-based implementation that enables its users to easily access and effectively manage the various data sets created throughout the development of the Project
- Provide a short description for each data set

The Zenodo platform offers the possibility to edit the metadata on the record. This feature enables edits for almost all of a record's metadata. From the upload type fields, it becomes evident that several different type of files such as publications, datasets, images, software's source code and others. This is demonstrated in Figure 7 below.

Edit upload

Instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Filename (1 files)	Size	Checksum ⓘ
amanda-project/ORDP-repository-v1.0.zip	271 Kb	md5:0405eff83bc6dc7c60febadbee0877dc


Note: File addition, removal or modification are not allowed after an upload has been published. This is because a Digital Object Identifier (DOI) is registered with [DataCite](#) for each upload. If you've made a mistake please [contact us](#).

If you have an updated version of your files, you can create a new record version, with a completely new DOI.

[New version](#)

Communities ⓘ recommended ▾

Start typing a community name...


Zenodo

Upload type required ▾

☐ Publication ☐ Poster ☐ Presentation ☒ Dataset ☐ Image ☐ Video/Audio ☐ Software ☐ Lesson ☐ Other

Figure 7 Edit field of Zenodo for metadata

A summary of the basic information that accompanies the file such as the digital object identifier, publication date, title, author, description, version, language, key words and additional notes are illustrated in Figure 8.

Basic information: required

- 1 **Digital Object Identifier**
 Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.
☒ Reserve DOI
- 2 **Publication date**
 Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.
- 3 **Title**
 Required.
- 4 **Authors**
 Optional.
- 5 **Description**

Sensor Measurements

 Required.
- 6 **Version** ☒
 Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended. See semver.org for more information on semantic versioning.
- 7 **Language**
 Optional. Primary language of the record. Start by typing the language's common name in English, or its ISO 639 code (two or three-letter code). See [ISO 639 language codes list](https://iso639.org) for more information.
- 8 **Keywords**
 Optional.
- 9 **Additional notes**
 Optional.

Figure 8 Basic information fields

Number	Field	Significance	Description
1	Digital object identifier	Optional	Identifier of the DOI
2	Publication date	Required	Date of first publication in case of upload file already published
3	Title	Required	Title of the upload file
4	Authors	Required	Author of the upload file
5	Description	Required	Description of the upload file
6	Version	Optional	Version of the upload file
7	Language	Optional	Primary language of the report
8	Keywords	Optional	Describe what the content about
9	Additional notes	Optional	Extra description for the upload file

Table 13 Basic information fields

The level of accessibility to the metadata can be chosen from the license tab. In addition, there is a field with regards to the grants that have funded the research, as shown in Figure 9.

License required

1 **Access right ***

- ☒ Open Access
- ☐ Embargoed Access
- ☐ Restricted Access
- ☐ Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

2 **License ***

Other (Open)

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the Other licenses available (Other (Open), Other (Attribution), etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org. If you think that a license is missing from the list, please contact us.

Funding recommended

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](https://openaire.eu/). Specify grants which have funded your research, and we will let your funding agency know!

3 **Grants**

European Commission (EU)

Start typing a grant number, name or abbreviation...

Optional. OpenAIRE-supported projects only. For other funding acknowledgements, please use the Additional Notes field.
Note: a human Zenodo curator will need to validate your upload - you may experience a delay before it is available in OpenAIRE.

[+ Add another grant](#)

Figure 9 License and Funding field

Number	Field	Significance	Description
1	Access right	Required	Accessibility of upload files
2	License	Required	License applies for the files
3	Grants	Optional	The grants that have funded the research

Table 14 License and Funding field

Number	Field	Significance	Description
1	Related/alternate identifiers	Recommended	Specify identifiers of related publications and datasets
2	Contributors	Optional	Names of contributors that help to create the upload file
3	References	Optional	References for the upload file
4	Journal	Optional	If the upload file has been published in journal fill out the basic information of journal such as title, volume, issue, pages
5	Conference	Optional	If the upload file has been published in conference fill out the basic information

			of conference such as title, acronym, dates, place, website, session, part
6	Book/report/chapter	Optional	If the upload file has been published in book form, fill out the basic information of the book such as publisher, place, ISBN, book title, pages
7	Thesis	Optional	If the upload file has been published as thesis, fill out the basic information of thesis such as awarding university, supervisors
8	Subjects	Optional	Add subject from a taxonomy or controlled vocabulary with term and the identifier

Table 15 Other Fields

All the metadata is stored internally in JSON-format according to a pre-defined JSON schema. They can be exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the OpenAIRE Guidelines).

3. Ethical concerns of the AMANDA project

The AMANDA consortium confirms that each partner will check with their national legislation/practice and their local ethics committee. That will provide guidelines on data protection and privacy issues, in terms of both data protection and research procedures in relation to any of the proposed public engagement and potential volunteer research activities. Any procedures for electronic data protection and privacy will conform to Directive (EU) 2016/680 and Regulation (EU) 2016/679 on the protection of personal data and its enactments in the national legislations.

The process of adhering to the applicable regulations begins with a thorough investigation of the EU and national research projects' ethical guidelines as well as the examination of the directives regarding privacy and protection of personal data and free movement of data issues. The legislation with which the AMANDA consortium must conform includes:

- The Universal Declaration of Human Rights
- The Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data
- The Directive 95/46/EC & Directive 2002/58/EC of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data
- Other European and national regulations related to privacy

The AMANDA project expects the development of a set of qualitative information collecting activities. In particular, interviews and questionnaires were planned, within Task 1.3. The double nature of consent appears again as both personal data and potentially sensitive information might be collected. Therefore, two issues become crucial from an ethical perspective: The confidentiality of the information and the anonymisation of personal data. The Code of Ethics of the International Sociological Association reminds researchers that "The security, anonymity and privacy of research subjects and informants should be respected rigorously" [4]. The sources of personal information obtained by researchers should be kept confidential, unless the informants have asked or agreed to be cited. Should informants be easily identifiable, researchers should remind them explicitly of the consequences that may follow from the publication of the research data and outcomes" [4]. From this article it is possible to extract some general rules that investigators must apply when designing and conducting their research:

- Information gathered from the participants should be kept confidential, unless specific consent to be cited is given by the participant.
- Information gathered should be anonymised and used only for the purpose for which it was collected.
- Participants must be informed when the investigator believes that some of the information shared may make them identifiable and the potential consequences.
- Participants must be given, in a clear and transparent manner, the opportunity to withdraw at any time and especially after being informed of their potential identification and potential the consequences.

In case the collected data contains personal information, data protection principles and legal requirements extracted from Regulation 2016/679 should be taken into consideration. In particular the investigator needs to put in practice organizational and technical measures directed to "minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing" [5].

In case the collected data contains personal information, the responsible partner of the AMANDA project should apply the following rules:

- Information collected from the participants should be anonymised. Responsible partners of the Consortium will prepare a summary, of the conducted research's results. The raw information will be kept in local resources by the partners under their own

responsibility and according to the data protection policies of their own organisations. Partners should pay special attention to the respect of the minimisation principle following article 89 (1) of Regulation 2016/679.

- Each task leader will collect the summaries and send them to the Ethics Helpdesk. The Ethics Helpdesk will review that no personal or sensitive information is contained in the summary, unless the participant has given specific consent. If needed, the Ethics Helpdesk can consult LEPPi during the process. The summary can be shared within the Consortium once this point is verified.
- The investigator must obtain specific consent from all the participants prior to their involvement in the different activities.
- The task leader of each of the activities will propose to the Ethics Helpdesk a text containing the specific information concerning the activity. The Ethics Helpdesk will validate the specific Informed Consent Form before it is used with any participants. Informed consent must be obtained, in written form.
- Oral informed consent is highly discouraged. Although oral consent is legally valid, the data controller must be able to "demonstrate that the data subject has consented to processing of his or her personal data" (Regulation 2016/679, article 7.1). Therefore, investigators should only use this procedure when there is no other possibility and after having consulted with the Ethics Helpdesk. The Ethical Body will evaluate the situation, bearing in mind the potential value of the information that could be obtained from the participant.

Duly signed Informed Consent forms, both written and electronic or proof of the oral consent, should be kept by the controller for a five years period to be available for auditing by the Ethics Helpdesk or any competent authority.

Miniaturizing the electronic system is going to be safe and non-intrusive. The project goals are technical. Most of the collected datasets are going to describe electronic system behaviour e.g. power consumption, voltage stability, radio connectivity performance and others. The measurements, which are not related to the technical evaluation of the system, are foreseen to be related to environmental conditions. An investigated subject is exposed to the measured conditions. There is no intention to directly gather measurement from bodies of living creatures. Thus, there is no interaction between the electronic system and the body of the subject. The project does not have the aim to perform human trials. However, experiments where the device is placed in the room where people are present is considered. The project is not in the scope of Utilisation of Genetic Resources the Access and Benefit Sharing (ABS).

4. Conclusions and future work

This report details the data management and ethics of the project. Section 2 provides insight on how the AMANDA Data Management Plan was created and implemented. This Section can later be used to track all the generated data in the project. It therefore satisfies the need of reusability of the generated data required by directive 2013/37/EU. Furthermore, the data which is generated during the project is mostly related to the performance of the electronic hardware. Also, this Section includes the description and systematisation of the data originated from the project. The data management approach presented in this report mostly consists of labelling and describing the generated data. In this way the data can be tracked during and after the project's execution.

During this project, no human trials are planned. However, data collection in form of questionnaires was conducted. Personal or sensitive data was labelled as such. Then, it was stored, analysed and used anonymously. The individuals were informed comprehensively about the intended use of the information collected from them. Participants provided their permission for data collection for a scientific purpose, with their active approval in form of a written consent. There is a potential for field tests, if time permits and lab testing is successful. These tests include the deployment of prototypes at the location of end users for a preliminary evaluation of the ASSC. However, this decision will be made towards the end of the project.

For each data set, ethical issues are considered separately in Section 3. The ethical aspect of each dataset is evaluated in the data description.

The data management plan & ethics document is an iterative report which will be updated again on M30 (v3) and M36 (v4). As a combined set of reports, it will document the progress of data generation, handling and storage.

Bibliography

- [1] K. Briney, "Data Management for Researchers: Organize, maintain and share your data for research success (Research Skills)," in *Data Management for Researchers*, 2015.
- [2] S. Brandt, "Data Management for Undergraduate Researchers: File Naming Conventions," Purdue University Libraries, [Online]. Available: <http://guides.lib.purdue.edu/c.php?g=353013&p=2378293>. [Accessed 24 06 2019].
- [3] D. Hillmann, "Dublin Core Metadata Initiative," 12 04 2001. [Online]. [Accessed 24 06 2019].
- [4] U. C. Faculty of Political Sciences and Sociology, "Code of Ethics," ISA Forum of Sociology International Sociological Association, [Online]. Available: <https://www.isa-sociology.org/en/about-isa/code-of-ethics>.
- [5] The European Parliament And The Council Of The European Union, "Regulation (Eu) 2016/679 Of The European Parliament And Of The Council," *Official Journal of the European Union*, vol. L, no. 119, pp. 1-88, 2016.