



Development of novel assessments for indoor air quality monitoring and impact on children's health

Data Management Plan v2

LEARN

AUTHOR SERAFIM KONTOS (ENVIROMETRICS)

Project details

Project Title	Development of Novel Assessments for indoor air quality monitoring and impact on children's health
Project Acronym	LEARN
Grant Agreement Number	101057510
Call	HORIZON-HLTH-2021-ENVHLTH-02
Starting date of the project	01/05/2022
Project Duration	48 months

The LEARN Consortium

Participants nº	Participant Organisation Name	Country
1 Coordinator	Laboratorio Iberico Internacional de Nanotecnología LIN (INL)	PT
2	Vrije Universiteit Brussel (VUB)	BE
3	Envirometrics Teknikoi Symvouloi Etaireia Periorismenis Efthynis (ENV)	EL
4	Katholieke Universiteit Leuven (KUL)	BE
5	Universiteit Hasselt (UH)	BE
6	Mann+Hummel GMBH (M&H)	DE
7	Nanotechnology Industries Association (NIA)	BE
8	Stichting IMEC Nederland (IMEC-NL)	NL
9	Aarhus Universitet (AU)	DK
10	F. Iniciativas España Mas i Mas SLU (FI)	ES
11	AlveoliX (AL)	CH

Project's aim

LEARNing about the status of air quality in schools and its impact on the cognition of children is a major cornerstone in LEARN. For that, we want to overcome the barriers of the currently existing technologies and take a bold step towards developing and deploying novel sensors to detect the presence of possibly harmful air pollutants such as volatile organic compounds and ultrafine particles.

We will measure and characterize indoor and outdoor air pollutants and evaluate the presence of biomarkers of exposure and their effect on children's cognition while trying to recapitulate those effects using *C.elegans* as a biosensor. Moreover, we will use advanced human-based in vitro models of lung and skin coupled with a revolutionary multi-sensing device to investigate their mechanisms of toxicity in real-time.

Novel remediation strategies will be explored to improve air quality and promote children's quality of life and life expectancy. For that, we will mobilize a group of eleven leading research



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

teams, unrivalled in their respective fields (environmental epidemiologists, toxicologists, air quality specialists, systems biology, engineers, and citizen/social scientists).

The scientific achievements expected to result from LEARN will unlock a large technology potential in IAQ for decades to come, leading to disruptive societal and economic impacts steaming from a radical improvement in the quality of life of children in Europe.

Project LEARN is part of the European indoor air quality and health cluster.

Document details

Deliverable Type	Report
Deliverable N°	D1.3
Deliverable Title	Data management Plan v2
Name of Lead Partners for this deliverable	Envirometrics Technikoι Symvouloi Etaireia Periorismenis Efthynis (ENV)
Version	Second version (2 nd version, M18)
Contractual delivery date	31/10/2023
Actual delivery date	31/10/2023
Dissemination Level	Public, fully open
Purpose	The current version of the LEARN Data Management Plan V2 describes in detail the data which will be produced from the project and the management of these according to FAIR principles and ethics.

Revision History

The following table describes the main changes done in the document since it was created

REVISION	DATE	DESCRIPTION	AUTHOR (ORGANIZATION)
V.0.1	31/10/2023	First draft of DMP v2	ENVIROMETRICS
V.0.2			
V.0.3			
V.0.4			
V.0.5			
V.0.6			
V.0.7			
V.0.8			
V.0.0			



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

Disclaimer

THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NON-INFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION, OR SAMPLE.

Any liability, including liability for infringement of any proprietary rights, relating to the use of information in this document is disclaimed. No license, express or implied, by estoppels or otherwise, to any intellectual property rights are granted herein. The members of the project Learn do not accept any liability for actions or omissions of Learn members or third parties and disclaim any obligation to enforce the use of this document. This document is subject to change without notice.

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. Neither the European Union nor the granting authority can be held responsible for them.



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

Executive Summary	6
1. Introduction	7
2. Data Summary	8
2.1. Types and formats of collected / generated data	8
2.2. Data utility	15
3. Fair data.....	16
3.1. Making data findable, including provisions for metadata	16
3.2. Making data accessible	17
3.3. Making data interoperable.....	19
3.4. Making data reusable.....	19
4. Other research outputs	21
5. Allocation of resources.....	22
5.1. Data management responsibilities	22
6. Data security.....	26
7. Ethics and other issues.....	27
8. Conclusions and next steps	28
Appendix: Detailed datasets description	29
A1: WP1 Datasets	29
A2: WP2 and WP3 Datasets	43
A3: WP4 Datasets	49
A4: WP5 Datasets	66
A5: WP6 Datasets	76
A6: WP7 Datasets	81
A7: WP8 Datasets	86

Executive Summary

This document constitutes the initial version of the Data Management Plan (DMP) in the framework of the LEARN project. LEARN aims to develop the next generation of sensors to detect Volatile Organic Compounds (VOCs) and Ultrafine Particles (UFPs) in order to identify potential risks and evaluate toxicity mechanisms. This process includes several activities of data gathering, production and/or processing with a view to generate meaningful perceptions that will feed into the project and lead to the successful development of the aforementioned sensors.

This deliverable constitutes the second version (v2) of LEARN's DMP, produced during the project and will serve as a living document. In this context, the second version of the project's DMP establishes the overall methodological principles referring to the management of data that will be gathered and/or produced during the LEARN project, protecting the ethical aspects of them. Moreover, it provides a detailed description of the data which will be produced from the project and the management of these according to FAIR principles and ethics.



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

1. Introduction

The current document represents the second version of the DMP for LEARN, which has received funding from the European Union's Framework Programme for Research and Innovation Horizon Europe under Grant Agreement No 101057510.

All participants of LEARN's consortium adhere to the data management principles to secure the meaningful data collection and processing during the project. The data are managed, archived and preserved in line with the structure and guidelines of the [Horizon Europe Data Management Plan Template](#).

According to this, DMP aims to achieve the following objectives:

- Describe the data management life cycle for the data to be gathered, generated, stored and/or reused in the framework of LEARN.
- Outline the methodology employed to secure the management of the data as well as to make them FAIR.
- Provide information on the project's data and the way in which it will be handled during and after the end of the project.
- Describe details on how the data will be made openly accessible and searchable to interested stakeholders as well as its curation and conservation.
- Present information on the resources to be allocated to make data FAIR clearly identifying responsibilities pertaining to data management, while addressing data security and ethical aspects.

With the above in mind, this second version of the DMP is structured in 8 distinct chapters, as follows:

Chapter 1 summarizes the structure and objectives of DMP.

Chapter 2 presents the collected/generated types of data of LEARN and their anticipated utility by selected stakeholders.

Chapter 3 outlines all the processes to make the data FAIR.

Chapter 4 includes the management of other research outputs that may be generated or re-used through the project.

Chapter 5 entails the responsibilities of each Work Package Leader (WPL) and Work Package Leader (WTL).

Chapter 6 describes the data security strategy in the framework of LEARN.

Chapter 7 addresses ethical aspects as well as other relevant issues pertaining to the data collected/generated or re-used during the implementation of the project.

Chapter 8 concludes on the project's data management plan.

An **Appendix** is also included at the end of the document, with tables providing details of the produced datasets from every WP.



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

2. Data Summary

LEARN will collect and generate meaningful data that do fall into special categories of personal data (personal information like age, gender etc., and health related information) as those are described within the General Data Protection Regulation (GDPR)¹. These data may be quantitative, qualitative or a blend of those in nature and will be analyzed from a range of methodological perspectives with a view to producing perceptions that will successfully feed LEARN'S activities, giving us the opportunity to deliver evidence-based results and ultimately achieve the objectives of the project.

2.1.Types and formats of collected / generated data

LEARN is set to collect and generate data of various structures and formats. There are two main aspects of the data:

- I. The process under which the underlying data are created/captured which includes electronic texts documents, spreadsheets, questionnaires and transcripts, among others.
- II. The storage format of the data in easily accessible formats like postscripts e.g., pdf, xps, etc.), machine readable formats (xml, html, etc.), spreadsheets, (e.g., xls, csv, etc.), text documents (e.g., docx, rtf, etc.), compressed formats (e.g., rar, zip, etc.) or any other format required by the objectives and methodology of the activity within the framework of which are produced.

In order to secure the interoperability and re-use of data special attention will be paid in using open formats² (such as csv, pdf, zip, etc.) where this is applicable, and / or machine-readable formats³ (such as xml, json, rdf, html, etc.). Furthermore, the data will be easily readable and freely usable in any software program employed by third parties interested in utilizing it.

The type and formats of the data collected/generated in the context of LEARN can be divided in two categories:

- I. Data collected through input methods (existing or created data) (Section 2.1.1)

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, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

² According to the [Open Data Handbook](#): "An open format is a file format with no restrictions, monetary or otherwise, placed upon its use and can be fully processed with at least one free/open-source software tool and it is not encumbered by any copyrights, patents, trademarks or other restrictions so that anyone may use it".

³ According to the [Open Data Handbook](#): "Machine readable formats are file formats that can be automatically read and processed by a computer. Machine-readable data must be structured data".



- II. Data collected/generated from dissemination, communication, stakeholder engagements, toolkits and clustering activities (Section 2.1.2)

2.1.1. Data collected through direct input methods

Direct input data are generated from literature reviews, experimental processes and interactions between consortium partners and external stakeholders. According to this, the latter undertake the role of a data subject that is a natural person whose personal data are being processed⁴. More specific, identification and selection of suitable data subjects depends on sampling in respect to the objectives of the activity for which the data are collected. The generated data can be divided into two main categories:

- **Quantitative data:** are numerical and gathered by counting or measuring⁵. For instance, the concentrations of chemical pollutants as well as particulate pollutants, classroom's environment (temperature, humidity, noise) etc. These data may be represented by numeric, interval and/or ratio scales.
- **Qualitative data:** are referred as categorical data, which can be arranged into categories based on physical traits, gender, colours or anything that does not have a number associated with it⁶. Other forms of qualitative data are written documents, interviews, and various forms of in-field observations (skillsets, country of origin etc.).

In the framework of the project several datasets will be produced with direct methods, which can be summarized in the following types:

Current status of children and their family background

The collection of these data includes some information about the physical measurements of children and the completion of questionnaires filled in by their parents for additional information about their background. Moreover, cognitive function tests will be performed to children to estimate the current cognitive level and related it to the Indoor Air Quality (IAQ) measurements. Some of them will be quantitative and some qualitative.

Data from classrooms' environment

The selected classrooms probably will include the volatile and semi-volatile compounds, which will be determined based on existing IAQ literature and multiple regulatory and health-based institutes. These measurements will be used for the development of a sensor selective to

⁴ 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, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>.

⁵ Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches. Boston: Pearson.

⁶ Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches. Boston: Pearson.



individual VOCs. In some classrooms, LEARN will perform stationary measurements on two locations in the classrooms. Additionally, a broad screening measurement will be performed in a limited selection of classrooms to detail the chemicals of interest. The above data will be mainly quantitative.

Current status of the IAQ

These data are mainly measurements or indicators of air quality in the classrooms such as concentrations of chemical pollutants (e.g., CO, NO₂, O₃, VOCs, PAHs, etc.) and particulate pollutants (e.g., PM_{2.5}, PM₁, UFPs, etc.) and parameters affecting it such as temperature, humidity, noise, light, ventilation system, window opening time, use of cleaning products etc. Furthermore, for an extra indication of the current IAQ status, urine samples from the children will be gathered. The collection of them will be performed in two steps: one before the implementation of air purification systems and one after it. Moreover, microorganisms in the filter will be collected in order to estimate the microbiological fraction present in the collected Particulate Matter (PM). The aforementioned data will be recorded in the form of spreadsheets, with some additional notes in document form linked to the lists.

Last but not least, subjective data will be collected through a short, seasonal survey with questions on 'compliments' or 'complaints' about the IAQ in order to be linked with the objective measurement data. These data will be gathered in a form of combination of qualitative and quantitative data.

Current status of the outdoor air quality

For the estimation of the outdoor air quality child's geocoded residence will be recorded in spreadsheet format. These data with children's addresses will be used for the modelling of the PM exposure at their homes. These data will be gathered in a form of combination of qualitative and quantitative data.

Research, design, development and prototype of innovative sensor technologies

The development of a sensor selective to individual VOCs and sensor system for UFPs, contains three main tasks, which generate a variety of data:

- A document of detailing the complete envisaged sensor system, its working principles and main design parameters in relation to the targeted sensitivity and detection range.
- Sensor system development, focusing on the engineering aspect of the sensor leading to sensor prototypes.
- Continuous testing and validation phase both in the lab and in the field to inform the status of sensor development and shape the next iterations of research and development.

The above results will be provided in a detailed document by IMEC-NL and INL, which includes information about the sensor's system development like the description of electron microscopy techniques, the development of the tuneable-wavelength (Vacuum Ultraviolet) VUV light source etc.

Evaluation of the toxicity

In a first-tier approach, traditional 2D cell culture models related to lung and skin models will be used for the evaluation of cellular death (cytotoxicity), oxidative stress, and proinflammatory effects. LEARN will take a step forward and work on the development of advanced in-vitro models of Lung-Blood-Barrier (LBoC) and Skin-on-Chip (SoC), which will be benchmarked



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

against the conventional mode and establish a testing protocol to evaluate the best on-chip administration method. These models will be described analytically in reports, which describe the above processes. The advanced lung and skin models will be used to assess the toxicological effects of chemical and biological air pollutants and to identify intra-, extra-, and inter-cellular biomarkers characteristics of air pollution-associated diseases. Finally, they will be compared with commercially available lung models.

Use of *C. elegans* as a biosensor

For further evaluation of the toxicity and neurotoxicity of the environment *C. elegans* will be used as an in-vivo model. The results will be translated/compared with the ones from the cognitive functional assessment (Work Package 2, WP2) by using the related transgenic/mutant organism at their L3/L4 larval stage. The panel of standardized new and realistic in vitro assays will be performed, matching the exposure scenario found in cohorts and preparing the ground for the definition of hazard profiles based on in silico testing. The protocol will be provided in a document format, following the standard approach to evaluate and propose items under the European Committee for Standardization (CEN).

Operation concepts for air purification systems based on UFPs concentration

In the framework of the LEARN project novel air purification systems will be developed to improve IAQ. Real IAQ data will be used to describe the efficiency of filters and air purifiers and efforts will be made to improve the performance of particle filters as well as adsorption layers. For the development of the above, the following processes will be developed:

- Demonstration of indoor air improvement, which includes the placement of air purifiers in classrooms based on Computational Fluid Dynamics (CFD) simulations.
- Filter media and adsorption material for the particle sizes and substances identified as important for IAQ.
- Modifications to the air purifier design and operation modes to improve the subjective IAQ level with specific focus on classrooms, e.g., noise, comfort flow rate, robustness, material selection.
- Test of samples and evaluation of potential health risks and evaluation of concentrations levels regarding health impacts to derive required filtration efficiency for given volume flow rate and contaminant concentration.

These processes will be written in analytical documents, which will also include quantitative results.

Organization of workshops

In the early stages of the project one FAIR data & ethical workshop will be organized when the use cases are clear and a second one before the final iteration of the innovation. These workshops will result in an overview of the FAIR data & ethical challenges within the project and possible mitigation actions. Through the implementation of the workshops, personal data from participants are going to be collected and recorded in participant lists format and spreadsheets filled with particulars. All the collected data will be aligned with GDPR.

Data collected from project coordination and management

During LEARN project, different activities will take place from which data will be created and handled. These can be from consortium meetings (name of attendees) to mailing lists and other day to day activities. The project team has prepared the Project Handbook (D1.1) where key



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

information on how partners will store and share data is defined. Data collected by partners through the implementation of workshops are more likely to be qualitative insights and will be recorded in various types of formats, with the photographs, videos, captured minutes, participant lists and document texts being the most expected.

AB feedback

During the implementation of LEARN project, there are several actions where guidance and feedback from experts is required. An AB composed of experts on the field is needed to provide necessary guidance and analysis of LEARN's activities and outcomes and to monitor, rate, and validate LEARN's support services and toolkit. Data collected via AB activities and involvement, such as discussions and digital validation workshops, are expected to be comments and proposals written in standard document texts and spreadsheets.

With the purpose of organizing the LEARN AB, personal data of AB members will be gathered. Collected data contain information such as name, surname, age, region, experience, education and expertise and will be saved in member lists. Finally, consent forms will be used for the ability of processing AB members' personal data and specific terms of reference will be developed to provide the basis for the activities of the AB.

Testing of the alert system

The alert system will be tested through a co-creation process which includes a citizen/user-centered approach through the organization of workshops and/or interviews with the stakeholders. The expected data will be in the form of documentation containing both quantitative and qualitative data. Both qualitative and quantitative data will be collected through interview questionnaires and guidelines during the first stage.

Policy insights set

In processes where desk research is included, there is an initial Policy insight set, which will be refined during the project. The generated data will be qualitative and written in a document/note form. The final Policy Insights is called "Policy Recommendations and Briefs" and will be refined according to European Union (EU) policy roundtable outcomes and comments as well as insights from other projects.

Insights from EU policy roundtable

An EU policy roundtable will be organized by partners and important data and views will be gathered from conversations during its implementation. The project's data will contribute to the scale up of the final set of "Policy Recommendations and briefs". Collected data will be mostly qualitative and will be written in document texts and participant lists.

2.1.2. Data collected/generated from dissemination, communication and clustering activities

Additional datasets are expected to be collected/generated from dissemination, communication and clustering activities. These can be summarized in the following main types:

Website analytics

In the framework of LEARN a website will be designed, which will contain a description of the project objectives, the proposed architecture, the consortium and other important up-to-date information. The website will use the managing environment of Wordpress with password-

protected administrative information and tools for updating the website, managing users' permissions and collecting website analytics.

The LEARN website will be supported by tracking and analytics software (Google Analytics and Google Search Console) for better understanding visitor's interactions with the website. In this way, the functionality of the tool will be improved and the visitor's privacy will be protected. To identify visitors on the website, cookies are used by Google Analytics to track their activity on the website navigation. Visitors will have the ability to delete or block website cookies. These cookies sometimes collect anonymous statistics about the visitor (such as gender, age, geographical location, and interests) and the data are stored by the analytics services that we use. These cookies also gather data regarding what pages they visit, and/or how long they stay on the page, and/or files they download. Tracking cookies from social media networks are also used, for customised advertising targeting visitors of the LEARN website on these platforms and to assess the performance of ads on these platforms. The data that these platforms collect is anonymised. That means that we cannot see the social media profiles of visitors.

At the same time, visitors can subscribe to the project's newsletter on the website. The newsletter mailing list is collected and managed through the Mailchimp software, integrated in WordPress. Newsletter subscribers will have the possibility to manage their subscription preferences by clicking on the link that will be available on each newsletter.

Project's email address

The project's official email address is hosted and managed through Gmail (Google Mail). Newsletters will be sent from this address using the Mailchimp software. Data is collected on Gmail from the messages that are sent to that email, for the purpose of providing an answer and analysing the stakeholder connections that are achieved.

Social Media statistics

A Twitter, LinkedIn and YouTube account for the LEARN project are created. The related data will be collected/generated through a periodic monitoring of the project's social media statistics (including Twitter, LinkedIn and YouTube) with a purpose to estimating and assessing the effectiveness and outcomes of the project's social media activity in terms of communication and dissemination. The data will be in qualitative and quantitative forms (e.g., followers, impressions, engagement, etc.). Moreover, these data will be used for a further analysis in order to improve the results so as to succeed the project's targets. In conclusion, the data will be stored in a Microsoft excel file (.xlsx) while at the same time the analysis of the results will be stored in a standard word document (.docx).

Data collected from project events

This type of data will be gathered during the project activities: (i) the different events (e.g., workshops, final event, etc.) organised by LEARN (either alone or jointly with other projects or initiatives) consisting of the participants lists and (ii) the participation of LEARN partners in relevant third-party events in order to reach out and engage stakeholders, thus collecting general information about the events attended and their outreach. The data will be both quantitative and qualitative in nature and will be stored in a standard spreadsheet (.xlsx).

Newsletter subscription (e.g., contact details of subscribers)

Project updates, results and achievements will periodically be disseminated in the form of a newsletter that will be spread via website, mail, project and institutional social media, etc. A subscription form in the project's website will facilitate the gathering of data in which any

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. Neither the European Union nor the granting authority can be held responsible for them.



Funded by
the European Union

interested stakeholder can voluntarily provide their contact details in a dedicated sign-up form, so as to receive the most up-to-date news and outcomes of the project. The data will be collected in order interested stakeholders can be informed about the LEARN as well as its digital Toolkits. Along these lines, the data will be comprised of a list of stakeholders along with their basic contact information: (i) email address, (ii) first and last name, (iii) country, and (iv) type of organization. A copy of this contact list is stored on MailChimp's (server which is used for e-mail campaigns and newsletter distribution. All personal information included in this contact list is used and protected according to MailChimp's Privacy Policy.

Furthermore, Mailchimp stores analytic data of the performance of the mailing campaigns that are launched, including email openings, clicks on buttons, time spent reading the email, etc.

When subscribing, visitors agree with the privacy policy and accept receiving the project's newsletters. At any time, they will have the possibility to manage their preferences or unsubscribe from the newsletters.

Data from dissemination and communication activities

This type of data concerns the data derived from the monitoring of the project's dissemination activities such as publications in journals, posts, etc. The data will consist of a list of publications and posts published by the consortium partners. In order to assess the efficiency of the dissemination activities, data will be collected in a template form. The data will relate with the recommended activities and will be saved in a single excel file.

Data collected for / generated by the use of LEARN toolkit

The LEARN will produce two different toolkits. The one is under T8.1 and will be developed to raise awareness about air contaminants commonly present at schools and their impact on the health and safety of children (in the long-term). The toolkit will also recommend remediation strategies to minimize children's exposure to these factors across multiple settings. The toolkit is expected to be used by schools in the consortium countries, with children and their parents as the primary target groups. This will be succeeded by providing detailed knowledge on the significance of continuous monitoring the IAQ and its consequences on the quality of life.

The second LEARN toolkit will be part of the Exploitation Strategy (Task 8.3) and will include exploitable results describing the way in which they can be exploited in a (non)-commercial sense after the end of the project, the target users and potential customers. According to the aforementioned functionalities the toolkit is expected to be used by pupils and their parents, schools, technology providers, supporting environment agents, research and academia and policy makers which in their turn will generate valuable data for consortium partners. The data will be both quantitative and qualitative and will be open source.

Furthermore, the toolkit's digital materials will be publicly available for downloads on the project's website. As for the offline materials (such as board games), the remaining units after the distribution to schools might be sold to the public if appropriate.

THE LIST OF THE DATASETS FROM EACH WP, WITH DETAILED PRESENTATION OF THE DATA TYPES, AND FORMATS CAN BE FOUND IN THE APPENDIX A, UNDER THE SECTION 1, "DATA SUMMARY".

2.2.Data utility

The stakeholders that may find the data to be collected/generated or re-used by the project useful are listed in the following table.

Table 1. Potential stakeholder's and data utility

Stakeholder Groups	Data Utility
School communities (teachers, students, parents)	The aim of LEARN project, is developing and deploying novel sensors to detect the presence of possibly harmful air pollutants such as VOCs and UFPs. These sensors will equip the selected classrooms and monitor the air quality for hazard air pollutants. Through the LEARN toolkit school communities can monitor the concentrations of air pollutants in classrooms and protect themselves in case these concentrations are dangerous for their health.
Policy makers at regional, national, and EU level related to public health	In the framework of LEARN, data will be collected related to the IAQ and will be correlated with the outdoor air quality and their impact on human health. Because of the sensors sensibility, the air pollutants monitoring will be more accurate and the results more trustworthy. Thus, policy makers can be alerted urgently about public health issues and be prepared properly.
Academic community in the field of public health	During the project innovative research technologies will be developed such as sensors, testing in biosensors etc. Open Access (OA) to scientific peer-reviewed publications will be ensured in the implementation of the LEARN Project. Gold OA costs are budgeted by the project partners, Green OA will be done through self-archiving published results, either on an institutional or disciplinary repository (Open-AIRE compliant). Research data will be available for access and re-use where applicable and under appropriate safeguards.
Project partners	The data collected/generated during the LEARN project will be necessary for project partners in order to produce evidence-based results and ultimately achieve the objectives of the project. At the same time, this data may be meaningful for project partners beyond the end of the project as well, enabling them to build and capitalise upon interesting ideas and opportunities that may emerge to ensure the long-term sustainability of the LEARN toolkit.

3. Fair data

The Data Management Plan guidelines from the Commission emphasize the importance of making the data generated by projects funded under Horizon Europe findable, accessible, interoperable as well as reusable, with a view to ensuring its sound management.

3.1. Making data findable, including provisions for metadata

3.1.1. Data discoverability and identification mechanisms

LEARN uses standard identification mechanisms in order to make the collected/gathered data findable and re-useable. More specifically, LEARN will assign globally resolvable Persistent Identifiers (PIDs) on open data, depending on the type of dataset (see Appendix A, Accessibility sections).

At this early stage of the project, besides the FAIR principles, the consortium has also identified the FAIR Digital Object (FDO) registry approach which enables the technical implementation of the FAIR principles to digital objects (such as code, models, workflows, services, etc.). The idea is to provide, for instance, explanations about interpreting and processing relevant digital artefacts, metadata standards, and validation rules for describing metadata.

In parallel, data that are not open will be stored in a searchable resource, i.e., the cloud web storage service of the project (MSTeams, Sharepoint). For storage, a name convention mechanism will be followed as described in section 3.1.2.

3.1.2. Naming conventions and versioning

Following a consistent set of naming conventions in the development of the project's data files can greatly enhance their searchability. With that in mind, LEARN creates consistent data file names that provide clues to their content, status and versioning, while also increasing their discoverability.

With that in mind and to facilitate the reference of the datasets that will be generated during its implementation, LEARN employs a standard naming convention that integrates versioning and takes into account the possibility of creating multiple datasets during an activity that entails data collection/generation.

In particular, the naming convention employed by the project is described below.

[Name of project] _ [Name of Study] _ [Number of dataset] _ [Issue Date] _ [Version number]

- **Name of project:** LEARN
- **Name of Study:** A short version of the name of the activity for which the dataset is created.
- **Number of dataset:** An indication of the number assigned to the dataset.
- **Issue Date:** The date on which the latest version of the dataset was modified (YYYY.MM.DD.).



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

- **Version number:** The versioning number of a dataset.

3.1.3. Metadata allowing discovery

In addition to compatible naming conventions and versioning, the project follows a metadata driven approach in order to allow further discovery and increase the searchability of the data. In simple lines, it is a tool that helps people discover, manage, describe, preserve, and build relationships with and between digital resources.

Data generated/used during LEARN are discoverable with metadata suitable to their content and format. The project employs metadata standards to generate rich and consistent metadata with a view to supporting the long-term discovery, use and integrity of the data.

3.1.4. Standards for metadata creation

LEARN employs standards for creating metadata for the project's collected and generated in order to improve their discoverability and searchability. Moreover, the applied metadata standards enable the integration of metadata from various sources into other technical systems.

LEARN's available open research data will potentially be published in Zenodo, depending on the dataset (see Appendix A, Accessibility sections). Their metadata will be exported in different standard and stored by Zenodo. Closed project data will also be interpreted as open and machine-readable metadata. The metadata to be used, will be compatible with Zenodo and in this way any initially closed data can potentially become open at the later stages of the project without metadata loss.

3.1.5. Search keywords

The data generated through the project will be provided with suitable key words to optimize their reuse by interested stakeholders. Keywords are used to add information to the gathered data and also to enable the easier description and interpretation of its content and value.

The choice of suitable keywords should follow the below principles:

- The who/what/when/where/why should be covered
- The consistency among the different keywords should be ensured
- The keywords should be relevant, clear and understandable

The keywords will be related to LEARN objectives (sensors, toxicity etc.). The keywords will describe the content of the datasets and avoid words used only once or twice within them.

3.2. Making data accessible

3.2.1. Repository

The data generated by LEARN and will be available for sharing and reuse and will be deposited to Zenodo (Section 3.1.4). Data in the Zenodo will be preserved long-term and the Digital Object Identifiers (DOIs) will function as perpetual links to the resources. They will remain



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

valuable since they are future proofed against Uniform Resource Locator (URL) or even protocol changes, through resolvers.

3.2.2. Data

LEARN, in line with the FAIR principles of data management in the context of Horizon Europe, adopts the good practice of making data open where this is possible and ensuring the proper protection of GDPR is followed. For this reason, LEARN asks from the partners to disseminate their data that potentially can be offered long-termly to external stakeholders and respect the confidentiality and privacy of stakeholders that contributed to the collection of this data.

Only anonymised and aggregated data will be made open to ensure that data subjects cannot be identified in any reports, publications and/or datasets resulting from the project. The relevant project partner in each case will undertake all the necessary procedures to anonymise the data in such a way that the data subject is no longer identifiable from public.

Data accessibility and availability

Public access to the open data will be made available and free of charge through Zenodo. The closed data will be stored and shared among authorized members of the consortium through cloud storage (i.e., MTeams) according to GDPR requirements.

Restrictions on use

As concern the project's open data, there are different levels of accessibility taking into account different issues (such as ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related, etc.).

The different levels of data accessibility are:

- **Open access:** These data are available for reuse (according to their accompanied license).
- **Embargoed status:** These data will be available after an "embargo" time period.
- **Restricted access:** These data will not be public and they will be shared only if the responsible partner allows it.
- **Closed access:** These data are accessible only by the members of the consortium.

A full description of the levels of accessibility for each dataset produced from each WP is presented in Appendix A, under Accessibility section.

3.2.3. Metadata

Metadata of deposited publications will be open under a **Creative Common Public Domain Dedication (CC 0)** or equivalent, in line with the FAIR principles for data management. Such metadata will include information about the following:

- The publication at hand (author(s), title, date of publication, publication venue);
- Reference to the Horizon Europe funding;
- The name of the project, including its acronym and Grant Agreement number;
- Any particular licensing terms which may apply (depending on the chose license);



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

- PIDs that have been attributed to the publication;
- Authors involved in the action, their organisations and the project itself.

Depending on the produced dataset, the metadata might also contain PIDs for any research outcome or any other tools and instruments needed to endorse the publication's conclusions. In case Zenodo stops its operation, all data, metadata, code and documentation uploaded will be transferred and stored to other available repositories. Meanwhile, the project's data with limited access will be deposited, along with their supplementary metadata, code and documentation (if exist), to the cloud web storage service employed by the project.

3.3. Making data interoperable

In order to succeed in this, LEARN uses metadata vocabularies, standards and methods that will increase the interoperability of the data collected/generated through its activities in its data management methodology.

The interoperability of the open-accessed data will be enabled through Zenodo containing community- endorsed practices, since its metadata are stored in JSON format according to a defined JSON schema.

3.4. Making data reusable

3.4.1. Documentation for validating data analysis and facilitating data re-use

To increase the reuse of the data according to FAIR principles, Zenodo follows the respective principles:

R1: (meta)data are richly described with a plurality of accurate and relevant attributes
Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.

- R1.1: (meta)data are released with a clear and accessible data usage license
License is one of the mandatory terms in Zenodo's metadata and is referring to an Open Definition license. Data downloaded by the users is subject to the license specified in the metadata by the uploader.

- R1.2: (meta)data are associated with detailed provenance
All data and metadata uploaded are traceable to a registered Zenodo user. Metadata can optionally describe the original authors of the published work.

- R1.3: (meta)data meet domain-relevant community standards
Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meet one of the broadest cross-domain standards available.



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

3.4.2. Availability for re-use

The re-use of data is a key component of LEARN's methodology for making data FAIR. It is expected that data will be available for re-use for some period after the end of its processing in the framework of the project.

3.4.3. Data provenance

Data provenance is the documentation of the data's origin as well as its producing methodology⁷. LEARN will use specific elements to record in a structured way the provenance of data.

LEARN metadata provenance will be provided through the DOI metadata record provided by Zenodo.

3.4.4. Data quality assurance processes

Quality Assurance (QA) and **Quality Control (QC)** are major parts of LEARN's DMP and are the first step before the publication of any data to Zenodo. They ensure the transparency, consistency, comparability, completeness and accuracy of the data.

QA is the review of the procedures, which are conducted without developing a dataset, by personnel not directly involved in the dataset development process⁸. In the framework of LEARN project, it takes the form of peer-reviews of methods and/or data summaries for the estimation of the dataset's quality and the identification of any need for improvement.

QC is a system for checking and maintaining the quality of the collected dataset⁹. The relevant procedures of LEARN are designed to implement technical checks measuring and controlling the data consistency, integrity, correctness and completeness as well as identifying and addressing errors and omissions. Some types of checks are:

- i) If there are transcription errors in data input
- ii) If scale measures are within the range of acceptable values
- iii) If proper naming conventions are used

⁷ <https://ardc.edu.au/resource/data-provenance/>

⁸ 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.

⁹ 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.



4. Other research outputs

It is expected that data will be available in the Data and Health Hub. During the implementation phase of LEARN project, data will be gathered from management and coordination activities. An Application Programming Interface (API) and a user interface will be developed, handling all data related functionalities. During the process of the design and implementation of the Air Quality and Health Data Hub, data will be collected for the formation of the data formats that will be handled by the Hub. The health-related data will be investigated to conform with the GDPR legislation and the chemical monitoring data in the Information Platform for Chemical Monitoring (IPCHEM) portal through the implementation of the API. The above data are expected to be both qualitative and quantitative and will be stored as written in standard document texts along with spreadsheets.



Funded by
the European Union

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. Neither the European Union nor the granting authority can be held responsible for them.

5. Allocation of resources

5.1. Data management responsibilities

For the effective, proper, and secure handling of the collected/generated data in the frame of LEARN, specific data management roles have been established within the data management methodology and procedures of the project. In this context, the following table represents the allocation of data management responsibilities among the members of the LEARN consortium per data collected/generated or re-used under each WP.

Table 2: Data management responsibilities of LEARN's partner per data collected/generated under each

WP	WPL	Data	Dataset	Tasks	Task Leader	Contributor(s)
WP1	INL	Data collected from project management and coordination	10	1.1	INL	All consortium
		AB Member list	11	1.1	INL	INL
		AB Feedback	11	1.1	INL	INL
WP2	UH	Questionnaires on personal data of children and their parents	1	2.1	UH	ENV, AU, VUB
		Results from five different cognitive function tests in school children	1	2.2	UH	AU
		Concentrations of VOCs and PM pollutants after the collection of environmental classroom's samples	2	2.3	UH	AU, M&H, KUL, INL
		Collection of urine and blood samples as an indicator of classroom IAQ	1	2.3	UH	AU, KUL, INL
		Measurements of classroom's temperature, light, humidity and noise	2	2.3	UH	AU, M&H, KUL, INL
		Measurements of school conditions (ventilation system, window opening time, use of cleaning products)	2	2.3	UH	AU, M&H, KUL, INL
		Records of children's geocoded residence	1	2.3	UH	AU, KUL, INL
		Collection of subjective data through a short, seasonal survey with questions on 'compliments' or 'complaints' about the IAQ in order to be linked with the objective	1, 3	2.4	VUB	AU, M&H, UH
WP3	KUL	Desk research based on existing IAQ literature and multiple regulatory and health-based institutes	3	3.1	KUL	AU, INL

		List of chemicals of interest in classrooms	2	3.1	KUL	AU, INL
		Actual measurements of VOCs, Aldehydes, PAHs in the classrooms	2	3.2	KUL	AU
		Actual measurements of all PM fractions	3	3.4	AU	KUL, INL, M&H
		Modelling of PM exposure at the home address of children	4	3.4	AU	KUL, INL,
		List, type and abundance of microorganisms in classrooms	3	3.5	INL	KUL
		Documentation of structural and chemical analysis of collected PM from the filters	3	3.6	INL	KUL
WP4	IMEC-NL	Analysed characteristics of envisaged sensor systems, description of working principles and main design	5	4.1	INL	IMEC-NL, M&H
		Description of VUV and its application on PID set-up	5	4.2	IMEC-NL	INL
		Quantitative measurements for the molecular ionization in a gaseous sample volume depending on the photon wavelength or equivalently the PID ionization energy	5	4.2	IMEC-NL	INL
		Comparison of UFPs performance with commercial needle-tip sources	5	4.2	IMEC-NL	INL
		Assessment of particle charging performance as function of voltage and air flow using controlled particle generation and reference monitors.	5	4.2	IMEC-NL	INL
		Description of NanomembraneS (NMBs)	5	4.3	INL	IMEC-NL
		Results of the production of photonic crystals suitable for Third Harmonic Generation (THG) at different wavelengths over the same wafer, after the successful demonstration of THG in SiO ₂ photonic crystal NMBs.	5	4.3	INL	IMEC-NL
		Results from the research of the designs for multiwavelength THG generation using a single nanostructure	5	4.3	INL	IMEC-NL
		Results from the Simulation of charged particle flow geared towards a low-cost PCB-based	5	4.3	INL	IMEC-NL

		sensor system design and considering the ionization source				
		Description of the design of Printed Circuit Board (PCB)-based electronics with selected components for low power, high resolution and current measurements.	5	4.3	INL	IMEC-NL
		Results from VOC sensor's testing after the use of single wavelength designs what will allow the determination and optimization of the main parameters of the VUV source: damage threshold, total VUV power generated and nanopatterned area needed.	5	4.3	INL	IMEC-NL
		Results from Testing the ultra-fine particle sensor prototype outside the lab related to its performance and robustness	5	4.4	IMEC-NL	AU, UH, KUL, INL
WP5	INL	Evaluation of cellular death, oxidative stress and proinflammatory effects of skin and lung cell lines using 2D cell cultures.	6	5.1	KUL	INL, NIA, AU, AL
		LBoC models for assessing the toxicological effects of PM and VOCs air pollutants	6	5.2	AL	INL, NIA, AU, KUL
		Chip design	6	5.2	AL	INL, NIA, AU, KUL
		Cell culture protocol developments	6	5.2	AL	INL, NIA, AU, KUL
		Protocol for the evaluation of the best on – chip administration methods.	6	5.2	AL	INL, NIA, AU, KUL
		Outcomes from the implementation of the testing protocols and qualified endpoints to assess the toxicity	6	5.3	INL	NIA
		Outcomes of the comparison of the analysis of LBoC models to commercially available lung models.	6	5.3	INL	NIA
		Skin – on – chip model	6	5.4	INL	NIA
		Protocol for using biosensors for air pollutants	7	5.5	INL	NIA, KUL, AU
WP6	M&H	Results/description of the placement of air purifiers in classrooms	8	6.1	M&H	UH, AU
		Description of filter media and adsorption material for the particle	8	6.2	M&H	KUL, IMEC-NL, INL

		sizes and substances identified as important for IAQ				
		Results/description of the modifications of air purifier design and operation modes in order to improve the IAQ level and adapt it to classrooms	8	6.2	M&H	KUL, IMEC-NL, INL
		Results from the test of samples and evaluation of potential health risks – Measurements of concentration levels regarding health impacts	8	6.3	M&H	KUL, INL
WP7	ENV	Collective data for the design of the data formats that will be handled by the Hub	9	7.1	ENV	All consortium
		Data integration plan and investigation of the ethics for the health-related data	9	7.1	ENV	All consortium
		User requirements for the functional and technical design of the Data API and Interface of the Data Hub	9	7.2	ENV	All consortium
		FAIR and ethics workshop (audio or video recording)	9	7.3	VUB	ENV, INL
		FAIR and ethics workshop (audio or video recording)	9	7.3	VUB	ENV, INL
		Chemical monitoring data	9	7.4	ENV	M&H
		Results from user testing of the alert system in the Data Hub. The test includes a citizen/user-centered approach through the organization of workshops and/or interviews with the stakeholders.	12	7.5	ENV	VUB, UH, M&H, INL
WP8	VUB	Website analytics	15	8.1	FI	All Consortium
		Social media statistics	17	8.1	FI	All Consortium
		Project events data	18	8.1	FI	All Consortium
		Project's email address	16	8.1	FI	All Consortium
		Newsletter subscriptions	19	8.1	FI	All Consortium
		Data from dissemination and communication activities	20	8.1	FI	All Consortium
		Participatory policy making	21	8.2	VUB	NIA, INL, UH, AU
		Exploitation Strategy and Sustainability Plan	21	8.3	NIA	All Consortium
		Policy recommendations	10	8.4	NIA	All Consortium
		IDEAL Cluster	10	8.5	FI	All Consortium

6. Data security

LEARN will handle any collected/generated or re-used data securely against any accidental loss and unauthorized access through the whole's project duration. In this way, the risk for a data breach is minimized. Especially, for cases of personal data collection/generation it is important that these data can be only accessible by those who have the respective authorization to do so.

Data collected/generated from partners should be processed using means (e.g., private servers) according to GDPR, which will ensure the data protection. Moreover, any necessary data security controls should be implemented in order to reduce any risk of information leaking/destruction to the minimum. In addition, the data will be backed up regularly taking into consideration the data's update and cruciality. The backed-up data will be saved in suitable storage media including external hard drives, flash drives, NAS devices and reputable cloud services in order to ensure its preservation, recovery and access. Lastly, frequent checks will take place for the data integrity.

Access to closed data will only be permitted to authorised project partners. In case there is a personal data breach, the responsible project partner will notify, without undue delay and, where feasible, not later than 72 hours after having become aware of it, its competent national supervisory authority (e.g., data protection authority) as well as the data subject(s) that may be affected by the breach. Moreover, the responsible partner will document any personal data breaches, including information such as the facts relevant to the breach, its effects and the remedial action(s) taken. Details for each dataset's protection can be found in Appendix A.

In the context of the project the identification and authentication access control are major components for the data's protection. In these lines, each project partner is responsible for enabling the application of appropriate access control to the data. For safety reasons, privacy policies will define the collection of online storage spaces, the processes and use of personal data, the security procedures followed, the user's rights as well as the employed policy about cookies.

7. Ethics and other issues

This section describes the ethical aspects of the LEARN DMP and the ethical consent of the collected/generated/reused data of the project. The project's data will be processed according to the relevant data protection legislation (e.g., GDPR). In accordance with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (GDPR), all personal data processed for project's activities will be:

- *processed lawfully, fairly and in a transparent manner in relation to the data subject;*
- *collected for specified, explicit and legitimate purposes relative to project's objectives and not further processed in a manner that is incompatible with those purposes;*
- *adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;*
- *accurate and, where necessary, kept up to date;*
- *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;*
- *processed in a manner that ensures appropriate security of the personal data.*

Additional details about the scope of activities that involve data processes (collection, generation, reuse) through the project along with the procedures for identifying the suitable stakeholders to participate in them should be defined by the respective WP Leaders. Furthermore, partners should process the data according to relevant EU and national regulations. Finally, it is not anticipated any data transfer outside the EU. In cases of data storage, the providers are committed to safeguard their compliance with the relevant GDPR requirements before start using their services.

Ethical challenges from the perspective of stakeholders will be collected during the workshops. These challenges are the result of a discussion about the positive and negative images of clearly defined pilot use cases and should lead to the determination of values that stakeholders find important (think of privacy, justice, applicability, reliability, solidarity, respect, autonomy). Next, action points to mitigate previously identified challenges on a technical, organisational and user level will be defined.

It is important to highlight that each partner is responsible for ensuring that the templates for any data gathering, like consents forms, are appropriately adjusted according to (i) the needs of the activity for which they are being used by them as well as to (ii) the relevant data protection laws and regulations applicable to their respective countries and / or organisation. All partners should keep records to demonstrate that data subjects have consented to the processing of their personal data and use consent management mechanisms that make it easy for individuals to withdraw their consent.

8. Conclusions and next steps

In this document the second version of LEARN's DMP has been presented. The types of the collected/generated data, their specifications, the protection and their compliance according with the FAIR principles were defined, as well the allocated resources and the ethics.

The DMP, as a living document, will be updated continually through the project, based on the latest outcomes and developments. As such, complementary information might be added, due to the complexity of collected/generated data. In addition, the two workshops planned in the frame of WP7 will engage relevant stakeholders and identify challenges concerning FAIR data and ethical principles (according to task 7.3), potentially providing additional feedback on the adoption of FAIR principles in the current document.

Appendix: Detailed datasets description

The appendix provides tables with the detailed information for each dataset expected to be produced by each WP, in the framework of the project. For each WP and dataset, a set of tables is provided with the following structure of eight (8) sections:

1. Dataset summary
2. Findability & Metadata
3. Accessibility
4. Interoperability
5. Re-use
6. Dataset financial management
7. Dataset protection
8. Dataset ethical and legal impact

A1: WP1 Datasets

In the frame of WP1, three (3) datasets will be produced:

- Advisory Board Contact List
- Consortium Contact List
- Data from Project management and Coordination

Below the Tables for each dataset and section are presented.

Advisory Board Contact List

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	No reuse of data
1.2 Types and formats of generated/re-used data	.xml
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	For the management and coordination of the Advisory Board (AB)
1.4 Expected size of the generated/re-used data	A few kb
1.5 Origin/provenance of the generated/re-used data	From the members of the AB
1.6 Data utility outside the project - To whom might your data be useful outside the project?	N/A

2. Findability & Metadata	
2.1 Data persistent identifier	NO

<i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content." https://guides.lib.uw.edu/research/PID</i>		
2.2 Metadata creation		NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>		N/A
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use		NO
2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	NO
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
	3.7 Directory and file naming convention will be used	NO

Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	NO
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	5 years
	3.11 Need of documentation/reference about any software to access or read the data	N/A
4. Interoperability		
4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Data won't be interoperable	
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A	
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	N/A	
5. Re-use		
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Data won't be reused	
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A	
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A	
Usability of generated in the project by third parties, in particular after the end of the project	N/A	

5.4 How data quality assurance will be achieved – what process will you follow to ensure that? N/A

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	PMs allocated to WP1 Overheads
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	Costs will be reported in the project budget (WP1)
6.3 How long they will be preserved after the end of the project	5 years by default

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
7.2 Data storage in trusted repositories for long term preservation and curation	NO
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation ¹⁰
7.5 Data need encryption	NO

¹⁰ https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	INL Data Protection Regulation ¹¹
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Storage within Microsoft Cloud, plus daily backup to INL's local storage.

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	Personal Data from the AB
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	N/A
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you	5 years

¹¹ https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

delete personal data that is no longer required to be stored?	
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	No protocols in place.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Joint Controllership Agreement in negotiation.

Consortium Contact List

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	No reuse of data
1.2 Types and formats of generated/re-used data	.xml
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	For project management and coordination
1.4 Expected size of the generated/re-used data	A few kb
1.5 Origin/provenance of the generated/re-used data	From the partners of the consortium
1.6 Data utility outside the project - To whom might your data be useful outside the project?	N/A

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content." https://guides.lib.uw.edu/research/PID</i>	NO
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist	N/A

in your discipline, please outline what type of metadata will be created and how.		
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use		NO
2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	NO
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
Metadata	3.7 Directory and file naming convention will be used	NO
	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	NO
	3.9 Metadata availability on information that enables the user to access the data	NO

3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	5 years
3.11 Need of documentation/reference about any software to access or read the data	N/A

4. Interoperability

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Data won't be interoperable
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	N/A

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Data won't be reused
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	N/A

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	PMs allocated to WP1 Overheads
---	-----------------------------------

6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	Costs will be reported in the project budget (WP1)
6.3 How long they will be preserved after the end of the project	5 years by default

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
7.2 Data storage in trusted repositories for long term preservation and curation	NO
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation ¹²
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	INL Data Protection Regulation ¹³
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Storage within Microsoft Cloud, plus daily backup to INL's local storage.

¹² https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

¹³ https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

8. Dataset ethical and legal impact	
8.1 Ethical or legal issues pertaining to the data	Personal Data from the Partners
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	N/A
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	5 years
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? For example: encryption of personal data in transit, anonymization of personal data locally	No protocols in place.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Joint Controllership Agreement in negotiation.

Data from Project Management and Coordination

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	No reuse of data
1.2 Types and formats of generated/re-used data	.xml
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	For the management and coordination of the Project (WP1)
1.4 Expected size of the generated/re-used data	A few kb
1.5 Origin/provenance of the generated/re-used data	Data from meetings, including, but not limited to participant list, recordings, minutes, presentations, reports, deliverables drafts and other documents. Data generated by the partners of the consortium
1.6 Data utility outside the project - To whom might your data be useful outside the project?	N/A

2. Findability & Metadata	
2.1 Data persistent identifier	NO
<i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	N/A
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO
2.5 Possibility of harvesting/indexing the metadata	NO
3. Accessibility	
Repository	3.1 Data deposition in trusted repository
	NO

	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
	3.7 Directory and file naming convention will be used	NO
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	NO
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	5 years
	3.11 Need of documentation/reference about any software to access or read the data	N/A
4. Interoperability		
	4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Data won't be interoperable

4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data (<i>e.g., other data from your project, or datasets from previous research</i>)	N/A
5. Re-use	
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Data won't be reused
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	N/A

6. Dataset financial management	
6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	PMs allocated to WP1 Overheads
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	Costs will be reported in the project budget (WP1)
6.3 How long they will be preserved after the end of the project	5 years by default

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
7.2 Data storage in trusted repositories for long term preservation and curation	NO
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation ¹⁴
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	INL Data Protection Regulation ¹⁵
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Storage within Microsoft Cloud, plus daily backup to INL's local storage.

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	Personal Data from the AB
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised

¹⁴ https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

¹⁵ https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

	or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	N/A
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	5 years
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	No protocols in place.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Joint Controllership Agreement in negotiation.

A2: WP2 and WP3 Datasets

In the frame of WP2 and WP3, one (1) unified dataset will be produced:

- Cohorts in schools and biomarkers of exposure

Below the Tables for each dataset and section are presented.

Cohorts in schools and biomarkers of exposure

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	No
1.2 Types and formats of generated/re-used data	Textual data: Questionnaires (e.g. education, occupation, diet, physical activity, stress) Numerical data: clinical (e.g. antropometrics, blood pressure) and experimental (e.g urinary black carbon, omics biomarkers, CANTAB) values
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	Data of WP2 and WP3 will be analysed in parallel
1.4 Expected size of the generated/re-used data	All the dataset generated will be less <100MB each. Arrays are expected to be <100GB.
1.5 Origin/provenance of the generated/re-used data	Questionnaires, biological matrices
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Other researchers with similar interests, consortia

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	YES
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	Metadata will be generated for questionnaires, in which explanation for each question and on data management of the answers will be given. Metadata for clinical parameters will include information of the clinical investigation troughs which the data were collected and data management of the data resulting from the investigation. Experimental data will be accompanied by metadata explaining the experiment and the data preprocessing (if any).

2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use		YES, cognition, school, children, indoor air quality, indoor pollution
2.5 Possibility of harvesting/indexing the metadata		YES
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	YES (sharepoint)
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO, as data include sensitive data
	3.4 Data accessibility through a free and standardized access protocol	YES
	3.5 Way to ascertain the identity of the person accessing the data	databases are accessible by other researchers upon formal requests after clearance from study PI's. Data will be accessible only when they will be completed, recorded and after successfully passed the data management processes. A DTA will be signed by the person accessing the data
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	YES, to approve access to sensitive data
	3.7 Directory and file naming convention will be used	YES
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	YES

3.9 Metadata availability on information that enables the user to access the data	YES
3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	25 years
3.11 Need of documentation/reference about any software to access or read the data	NO

4. Interoperability

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	We will use standard format for data such (eg. for epigenetics Genomicratio set will be generated, for urinary black carbon we will generate datasets expressing BC black carbon per milliliter urine)
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	NO

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Readme files with information on methodology and codebooks will be provided
5.2 Free availability of the data in the public domain to permit the widest re-use possible	NO
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	NO
Usability of generated in the project by third parties, in particular after the end of the project	NO
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	Data will be checked for quality via automated processes (eg. histograms for checking for outliers) and through manual checks

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	Direct and indirect costs related to data management, storage, archiving, re-use and security
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	Capacity and infrastructure already in place.
6.3 How long they will be preserved after the end of the project	25 years

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Sharepoint and U Hasselt institutional backup-system will be used
7.2 Data storage in trusted repositories for long term preservation and curation	NO
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	U Hasselt has a license agreement whereby the storage of the data is well protected
7.5 Data need encryption	YES, personal data (names, surnames, dates, addresses)
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES, the data protection officer of the research centre of Universiteit Hasselt, Martelarenlaan 42, B-3500, e-mail: DPO@uhasselt.be will be responsible for data protection
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Automatic online backup will be performed and the research team will make a back-up on an external drive

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
--	----

8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	YES
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	Use of pseudonymized data
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	YES, pseudonymisation
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	After 5 years from the end of the study personal data will be deleted.
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? For example: encryption of personal data in transit, anonymization of personal data locally	Personal data will not be shared.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	All the partners from the project will sign a Data Joint Controllership Agreement in order to ensure that any personal data which needs to be shared among partners for the purposes of the project are protected.

A3: WP4 Datasets

In the frame of WP4, three (3) datasets will be produced:

- PM benchmarking data collected in workplace vitality HUB
- VOC's Prototype Validation in Field Environment
- VOC's Prototype Validation in Laboratory Environment

Below the Tables for each dataset and section are presented.

PM benchmarking data collected in workplace vitality HUB

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Time series of newly developed PM sensors against reference sensors, in CSV format or similar. The data will be combined with environmental information such as temperature, humidity, occupancy, etc..
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	Part of WP4, Task 4.4 Sensor Testing and Validation
1.4 Expected size of the generated/re-used data	Assuming a sampling interval of several minutes, the expected size is less than 1e6 samples.
1.5 Origin/provenance of the generated/re-used data	The data will be uniquely generated from an experiment, executed at the workplace vitality HUB in Eindhoven (Netherlands), in which newly developed PM sensors will be compared to reference devices.
1.6 Data utility outside the project - To whom might your data be useful outside the project?	- The data might be relevant for PM sensor developers / manufacturers, as it will contain information on expected concentrations in a normal office environment - The data might also be useful for researchers in the field of human exposure to air pollution, as it might provide data regarding exposure of office workers.

2. Findability & Metadata	
2.1 Data persistent identifier	YES/NO (If yes, please provide it)

	<i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
	2.2 Metadata creation	NO
	2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	N/A
	2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	N/A
	2.5 Possibility of harvesting/indexing the metadata	N/A
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	YES/NO (If yes, <u>state the repository</u> and if you have explored the <u>appropriate arrangements</u> for the deposition)
		YES, assuming that the repository will be provided for by the project
	3.2 Identifier assigned to data by the repository	YES/NO (If yes, state if the repository will resolve the identifier to a digital object) NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	YES/NO (If the dataset cannot be shared <i>(or need to be shared under restricted access conditions)</i> , <u>explain why</u> , clearly separating legal and contractual reasons from intentional restrictions.) YES
	3.4 Data accessibility through a free and standardized access protocol	YES/NO (If there are restrictions on use, state how access will be provided to the data, both during and after the end of the project) YES
	3.5 Way to ascertain the identity of the person accessing the data	NO
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)	NO

	3.7 Directory and file naming convention will be used	YES, assuming that this will be guided by the project
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	YES/NO (If no, please <u>clarify why</u>) NO, creating metadata not foreseen (see 2.2)
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	t.b.d.
	3.11 Need of documentation/reference about any software to access or read the data	YES/NO (If yes, please state if it will be possible to include the relevant software (e.g., in open-source code)) NO
4. Interoperability		
	4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Data will be shared in CSV format, stored in SI units
	4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	YES/NO (Please state if you will openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them) NO
	4.5 Qualified references to other data included in the data (e.g., other data from your project, or datasets from previous research)	YES/NO (If yes, please state the references) NO
5. Re-use		
	5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Readme file (.txt format)
	5.2 Free availability of the data in the public domain to permit the widest re-use possible	YES/NO

	YES, if access provided by the project AND unless limited by the supplier of the sensors
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	YES/NO, if yes please state how the data will be licensed for re-use NO
Usability of generated in the project by third parties, in particular after the end of the project	YES/NO Yes
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	t.b.d.

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	Under the assumption that tools will be provided by the project to publish the data, there are no additional costs expected
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	
6.3 How long they will be preserved after the end of the project	

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Not required
7.2 Data storage in trusted repositories for long term preservation and curation	YES/NO, please state for how long and which data YES, if provided by the project
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	YES/NO YES, this is completely relying on third-party services
7.4 Procedure(s) in place to follow in the case of a data breach	No personal data will be collected
7.5 Data need encryption	YES/NO (if yes, please state for which data and the process)

7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	NO YES/NO (if yes, please state what guidelines/policies are in place) DPO, but involvement is not expected to be required for this project
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	NO

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	None
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? 'Pseudonymisation' of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	NO – no personal data will be collected
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why?	N/A

For example: encryption of personal data in transit, anonymization of personal data locally	
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A

VOC's Prototype Validation in Field Environment

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	YES Re-use data (DATA 1) from the VOC's sensor prototype validation in laboratory environment
1.2 Types and formats of generated/re-used data	Two types of data: DATA 1 - Validation in the field laboratory of the sensitivity and selectivity of the sensor prototype for detecting the selected VOC's with respect to reference instrumentation Quantitative data numerical data will be generated with a software named LabView. This software generates TDMS files which are binary files. Proprietary format of binary files produced by National Instruments. If necessary these TDMS files can be converted in xml type files and accessed in open source programs. DATA 2 - Video demonstration of the sensor prototype in field environment related to MS8
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	WP4 – Volatile Organic Compound (VOC) sensor prototype development
1.4 Expected size of the generated/re-used data	GB due to DATA2
1.5 Origin/provenance of the generated/re-used data	DATA 1 - Quantitative data generated from the sensor prototype measurement using LabView software

	DATA 2 - Video demonstration of the VOC sensor in field environment
1.6 Data utility outside the project - To whom might your data be useful outside the project?	To whom might be interested in Indoor Air Quality in particular, selective VOC detection using the technology under development – Sample Ionization Potential Spectroscopy

2. Findability & Metadata

2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	NO The metadata created are the metadata related to the software used in the measurements of the sensor (LABVIEW) and the metadata to the video format of the file.
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	NO
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO
2.5 Possibility of harvesting/indexing the metadata	NO

3. Accessibility

Repository	3.1 Data deposition in trusted repository	NO Regarding the data of the VOC sensor if successful we will share it through an open access publication. For the video, we think it does not makes sense to place it in a repository.
	3.2 Identifier assigned to data by the repository	NO

Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	DATA 1 - Quantitative data obtained with the sensor will be open access if accepted manuscript publication
	3.4 Data accessibility through a free and standardized access protocol	Both type of data mentioned in 1.2 will be accessible in the WP4 folder created in MStTeams belonging to the Microsoft cloud.
	3.5 Way to ascertain the identity of the person accessing the data	We are not aware how we can ascertain the identity of the person accessing the data in MStTeams
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
	3.7 Directory and file naming convention will be used	[name of the project]_[name of the study]_[number of dataset]_[issue date]_[version number] Store in WP4 in MStTeams sharepoint LEARN Project/WP4/Files
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	NO Metadata related to open access publications will be open Metadata of video demonstration will not be open access
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	5 years by default

3.11 Need of documentation/reference about any software to access or read the data	NO We generate data from the sensor and the video. The data from the sensor can be converted to be read in open source related excel programs. For the video any related video program can open the file
4. Interoperability	
4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Semantical interoperability: Non-numerical data will be provided in English. Video demonstration will have legends and narration in English
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	We will create the ontology related to the method of detection which is: Sample Ionization Potential Spectroscopy (SIPS) which has been used in project proposal and to our best knowledge has not been used in any scientific context before
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	YES Data contained in documents related to WP 4 - VOC's Prototype Validation in Laboratory Environment
5. Re-use	
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Readme file related to the quantitative data obtained – operation of the sensor prototype
5.2 Free availability of the data in the public domain to permit the widest re-use possible	Data is available once a scientific manuscript is accepted for an open access publication
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	NO
Usability of generated in the project by third parties, in particular after the end of the project	NO
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	To our best knowledge not defined for WP4

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	PMs allocated to WP4 Overheads
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	Costs will be reported in the project budget
6.3 How long they will be preserved after the end of the project	5 years by default

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
7.2 Data storage in trusted repositories for long term preservation and curation	NO
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES Storage at Microsoft Cloud, plus daily backup to INL's local storage.
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Storage at Microsoft Cloud, plus daily backup to INL's local storage.

8. Dataset ethical and legal impact	
8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	Intellectual Property will be handled by the IP Exploitation and Knowledge Transfer unit at INL. This is related to MS7 described in project – VOC sensor prototype in the laboratory TRL 1/2
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully anonymised data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	NO
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A. We will not collect personal data under WP4
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? For example: encryption of personal data in transit, anonymization of personal data locally	N/A. We will not collect/transfer/share personal data under WP4
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A. We will not collect/transfer/share personal data under WP4

VOC's Prototype Validation in Laboratory Environment

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Two types of data: 3. DATA 1- Validation in the laboratory of the sensitivity and selectivity of the sensor prototype for detecting the selected VOC's with respect to reference instrumentation Quantitative data – numerical data will be generated with a software named: LabView. This software generates TDMS files which are binary files. Proprietary format of binary files produced by National Instruments 4. DATA 2- Video demonstration of the sensor prototype in laboratory environment related to MS7
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	WP4 – Volatile Organic Compound (VOC) prototype sensor development
1.4 Expected size of the generated/re-used data	GB due to DATA2
1.5 Origin/provenance of the generated/re-used data	DATA 1 - Quantitative data generated from the sensor prototype measurement using LabView software DATA 2 - Video demonstration of the VOC sensor in laboratory environment
1.6 Data utility outside the project - To whom might your data be useful outside the project?	To whom might be interested in selective VOC detection using the technology under development – Sample Ionization Potential Spectroscopy

2. Findability & Metadata

2.1 Data persistent identifier	NO
<i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of</i>	

the metadata or content." https://guides.lib.uw.edu/research/PID		
2.2 Metadata creation		YES for the quantitative data described in 1.2 and the video demonstration
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>		Every TDMS file contains two types of data: meta data and raw data. Meta data is descriptive data stored in objects or properties. Data arrays attached to channel objects are referred to as raw data. Meta Data —Contains names and properties of all objects in the segment. For objects that include raw data (channels), the meta data part also contains index information that is used to locate the raw data for this object in the segment.
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use		NO
2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	NO
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	Research data will be open access only through a manuscript publication or deliverables for public dissemination beyond WP4
	3.4 Data accessibility through a free and standardized access protocol	Both type of data mentioned in 1.2 will be accessible in the WP4 folder created in MSTEams belonging to the Microsoft cloud.
	3.5 Way to ascertain the identity of the person accessing the data	We are not aware how we can ascertain the identity of the person accessing the data in MSTEams

Metadata	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
	3.7 Directory and file naming convention will be used	[name of the project]_[name of the study]_[number of dataset]_[issue date]_[version number] Store in WP4 in MStTeams sharepoint LEARN Project/WP4/Files
	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	NO Metadata related to open access publications will be open Metadata of video demonstration will not be open access
	3.9 Metadata availability on information that enables the user to access the data	NO
Metadata	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	5 years by default
	3.11 Need of documentation/reference about any software to access or read the data	NO We generate data from the sensor and the video. The data from the sensor can be converted to be read in open source related excel programs. For the video any related video program can open the file
4. Interoperability		
4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Semantical interoperability: Non-numerical data will be provided in English. Video demonstration will have legends and narration in English	
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	We will create the ontology related to the method of detection which is: Sample Ionization Potential Spectroscopy (SIPS) which has been	

	used in project proposal and to our best knowledge has not been used in any scientific context before
4.5 Qualified references to other data included in the data (<i>e.g., other data from your project, or datasets from previous research</i>)	NO
5. Re-use	
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Readme file related to the quantitative data obtained – operation of the sensor prototype DATA 2 - The video shall not be re-used
5.2 Free availability of the data in the public domain to permit the widest re-use possible	NO Availability of the data only if a publication is accepted
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	NO
Usability of generated in the project by third parties, in particular after the end of the project	NO
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	To our best knowledge not defined for WP4

6. Dataset financial management	
6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	PMs allocated to WP4 Overheads
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	Costs will be reported in the project budget
6.3 How long they will be preserved after the end of the project	5 years by default

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
7.2 Data storage in trusted repositories for long term preservation and curation	NO
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES Storage at Microsoft Cloud, plus daily backup to INL's local storage.
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Storage at Microsoft Cloud, plus daily backup to INL's local storage.

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	Intellectual Property will be handled by the IP Exploitation and Knowledge Transfer unit at INL.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be	NO

<i>used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully ‘anonymised’ data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/</i>	
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A. We will not collect personal data under WP4
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	N/A. We will not collect/transfer/share personal data under WP4
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A. We will not collect/transfer/share personal data under WP4

A4: WP5 Datasets

In the frame of WP5, two (2) datasets will be produced:

- Development of multi-sensing skin-on-chip
- Testing *C. elegans* as biosensors for toxicological effects

Below the Tables for each dataset and section are presented.

Development of multi-sensing skin-on-chip

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	YES, SOPs available from other EU projects will be employed.
1.2 Types and formats of generated/re-used data	(pdf, xps, etc.), machine readable formats (xml, html, etc.), spreadsheets, (e.g., xlsx, csv, etc.), text documents (e.g., docx, rtf, etc.), compressed formats (e.g., rar, zip, etc.)
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	Task5.4- we will evaluate the effects of tested air pollutants on the skin-on-chip model and compare 2D commercially available models. Data from these tests will be generated and stored.
1.4 Expected size of the generated/re-used data	Many GB
1.5 Origin/provenance of the generated/re-used data	In house
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Other EU projects, regulators, industry

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	YES/NO (If yes, please provide it) Not know yet.
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not	N/A

exist in your discipline, please outline what type of metadata will be created and how.		
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use		NO
2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	YES, Zenodo (if project decides to use this)
	3.2 Identifier assigned to data by the repository	YES/NO (If yes, state if the repository will resolve the identifier to a digital object) If Zenodo used, then an identifier is assigned to the data.
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	<i>At least at the end of the project data will be made open access after publication and IP protection.</i>
	3.4 Data accessibility through a free and standardized access protocol	Data will be shared with partners of the consortium.
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
Metadata	3.7 Directory and file naming convention will be used	ProjectName_StudyName_DatasetNumber_Issue date_Version number
	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means</i>	NO. No metadata will be purposely created, except metadata automatically generated by the different types of files.

<i>relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	
3.9 Metadata availability on information that enables the user to access the data	NO
3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	N/A
3.11 Need of documentation/reference about any software to access or read the data	NO
4. Interoperability	
4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	SOPs, ISO and OECD guidelines will be employed
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data (e.g., other data from your project, or datasets from previous research)	YES/NO (If yes, please state the references) Not known yet
5. Re-use	
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Data units are clearly provided in the files. SOPs will be publicly available after IP protection and publication.
5.2 Free availability of the data in the public domain to permit the widest re-use possible	YES
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	
Usability of generated in the project by third parties, in particular after the end of the project	YES

5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	Quality control will be obtained by specific controls (neg. and post.), and equipment will be calibrated.
--	---

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	LEARN costs will cover storage, security, accessibility, and licensing.
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	WP5 budget
6.3 How long they will be preserved after the end of the project	4 -5 years

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has a private data protection system.
7.2 Data storage in trusted repositories for long term preservation and curation	Data from qualified endpoints to assess toxicity will be preserved in Zenodo.
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES, INL regulation
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	INL specific storage system

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	N/A
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	N/A
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	N/A
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? For example: encryption of personal data in transit, anonymization of personal data locally	N/A
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A

Testing C. elegans as biosensors for toxicological effects

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO

1.2 Types and formats of generated/re-used data	Docx; xlsx; pdf; pptx; png; tiff; jpeg
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	In WP5 (Task 5.5), we will use <i>C. elegans</i> as a biosensor to evaluate toxicity of air pollutants. Data will be generated related to this experiments.
1.4 Expected size of the generated/re-used data	Several GB
1.5 Origin/provenance of the generated/re-used data	Data from the implementation of the testing protocols and qualified endpoints to assess toxicity and comparison between models.
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Toxicologists, communities working with Non-animal models and new approach methodologies (NAMs), regulators,

2. Findability & Metadata

2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	YES/NO (If yes, please provide it)
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	N/A
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	N/A
2.5 Possibility of harvesting/indexing the metadata	NO

3. Accessibility

Repository	3.1 Data deposition in trusted repository YES Before publication, data will be stored at the INL Onedrive. After that, and once no IP issues apply, data will be made public and stored in Zenodo.
-------------------	--

	3.2 Identifier assigned to data by the repository	YES/NO (If yes, state if the repository will resolve the identifier to a digital object)
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	YES/NO (If the dataset cannot be shared <i>(or need to be shared under restricted access conditions)</i> , <u>explain why</u> , clearly separating legal and contractual reasons from intentional restrictions.) Data will be made open access after publication and IP protection (if applicable), most likely at the end of the project (2026).
	3.4 Data accessibility through a free and standardized access protocol	YES/NO (If there are restrictions on use, state how access will be provided to the data, both during and after the end of the project) Data will be accessible via sharepoint for interested partners involved in the task/WP.
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
	3.7 Directory and file naming convention will be used	ProjectName_StudyName_DatasetNumber_Issue date_Version number
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	NO No metadata will be purposely created, except metadata automatically generated by the different types of files.
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	N/A
	3.11 Need of documentation/reference about any software to access or read the data	NO

4. Interoperability

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Data will be described following the OECD Guidance Document 211 for non-guideline in vitro methods, whenever possible.
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	YES/NO (If yes, please state the references)

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Data units are clearly provided in the files. SOPs will be publicly available after IP protection and publication.
5.2 Free availability of the data in the public domain to permit the widest re-use possible	YES
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	YES/NO, if yes please state how the data will be licensed for re-use
Usability of generated in the project by third parties, in particular after the end of the project	YES
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	Peer-review of methodologies and dataset, adequate team training. Quality checks will be frequently performed (using positive and/or negative controls) and equipment will be periodically calibrated to ensure quality control.

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	Expected costs are related to storage, security, accessibility, and licensing.
6.2 How they will be covered	Costs will be covered by WP5 budget

<i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	
6.3 How long they will be preserved after the end of the project	5 years by default

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	INL has in place appropriate technical and organisational measures to protect against unauthorised or unlawful processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data, appropriate to the harm that might result from the unauthorised or unlawful processing or accidental loss, destruction or damage and the nature of the data to be protected, having regard to the state of technological development and the cost of implementing any measures.
7.2 Data storage in trusted repositories for long term preservation and curation	YES, please state for how long and which data Data from qualified endpoints to assess toxicity will be preserved in Zenodo.
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Art. 13 of INL Data Protection Regulation ¹⁶
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES INL Data Protection Regulation ¹
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Storage within INL's local storage.

8. Dataset ethical and legal impact

¹⁶ https://inl.int/wp-content/uploads/2020/04/INL-Data-Protection-Regulation_REGDIR003.pdf

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	N/A
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	N/A
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	N/A
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? For example: encryption of personal data in transit, anonymization of personal data locally	N/A
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A

A5: WP6 Datasets

In the frame of WP6, one (1) dataset will be produced:

- CFD Simulations

Below the Tables for the dataset and each section are presented.

CFD Simulations

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	CAD and simulation data in application specific format (STL, msh, ANSYS result files)
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	CFD simulations for optimized filter unit positioning to increase efficiency of air purification
1.4 Expected size of the generated/re-used data	Several 100 GB
1.5 Origin/provenance of the generated/re-used data	CAD system / simulation software (ANSYS)
1.6 Data utility outside the project - To whom might your data be useful outside the project?	IDEAL cluster partners Schools

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	YES
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	NO
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO

2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	Only in project sharepoint
	3.2 Identifier assigned to data by the repository	N/A
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	No, as data shows proprietary information owned and developed by MANN+HUMMEL GmbH
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	Data only available to those granted access internally.
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	NO
Metadata	3.7 Directory and file naming convention will be used	YES
	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	YES
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	1 year
	3.11 Need of documentation/reference about any software to access or read the data	YES (ANSYS is commercial software, no open-source alternative available)
4. Interoperability		

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Internal methodologies and formats will be utilized.
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	NO
5. Re-use	
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	Report on methodology and results
5.2 Free availability of the data in the public domain to permit the widest re-use possible	NO
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	Not yet known
Usability of generated data in the project by third parties, in particular after the end of the project	NO
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	3D model of chosen classrooms carried out using LIDAR scanner, then uploaded into CAD model and CFD simulations carried out on reputable software by experienced modeller

6. Dataset financial management	
6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	None
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	N/A

6.3 How long they will be preserved after the end of the project	10+ years
--	-----------

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Using established PLM system CIM Database
7.2 Data storage in trusted repositories for long term preservation and curation	YES (10 years internal data storage in PLM system, CAD and simulation data in application specific format (STL, msh, ANSYS result files))
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	YES
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Data to be stored on server and later archived in long-term database.

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	N/A
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	N/A
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category?	N/A

Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	N/A
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A

A6: WP7 Datasets

In the frame of WP7, one (1) dataset will be produced:

- Data related to the FAIR and ethics workshop

Below the Tables for the dataset and each section are presented.

Data related to the FAIR and ethics workshop

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Personal data: audio and/or video recording
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	WP7 /T7.3 FAIR Data principles adaptation and ethics
1.4 Expected size of the generated/re-used data	Personal data: audio and/ or video recordings of 2,5 hours each
1.5 Origin/provenance of the generated/re-used data	Audio and/ or video recording
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Nobody

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	N/A
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO

2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	YES (Sharepoint)
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO The personal data will not be shared, restricted access
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	N/A
	3.9 Metadata availability on information that enables the user to access the data	NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	N/A
	3.11 Need of documentation/reference about any software to access or read the data	N/A

4. Interoperability

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	N/A
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	N/A

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	N/A
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	N/A

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	N/A
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	N/A
6.3 How long they will be preserved after the end of the project	N/A

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Storage on Sharepoint LEARN and VUB
7.2 Data storage in trusted repositories for long term preservation and curation	Until 5 years after the project ending
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Inform the data subject and delete the personal data from the database All LEARN partners will sign a Joint Controllorship Agreement in order to be compliant with GDPR legislation.
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES, VUB policies All LEARN partners will sign a Joint Controllorship Agreement in order to be compliant with GDPR legislation.
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Sharepoint

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	YES
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	Only access by the partners involved in the task Anonimisation of the transcription.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? 'Pseudonymisation' of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other	Yes

words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	5 years after the project ending Deletion of the file
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	None
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	All LEARN partners will sign a Joint Controllership Agreement to be compliant with GDPR legislation.

A7: WP8 Datasets

In the frame of WP8, six (6) datasets will be produced:

- Contact details for sending email communications about the project
- Performance analytics of online communication tools (website, social media and newsletters)
- Contact details of workshop participants and final conference events
- Data related to the participatory policy making processes
- Research data related to the participatory policy making processes
- Subjective indoor air quality

Below the Tables for each dataset and section are presented.

Contact details for sending email communications about the project

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Contact details of newsletter subscribers (people who agree to receive the project's email communications): email address and first name. Data is stored in an online database on the MailChimp platform, and also in CVS/XLS files in FI Group's servers.
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	This data will be used for T8.1 under WP8. In particular, for the newsletters of the project, as stated in section 3.2.4 of LEARN's Communication & Dissemination Plan (D8.1).
1.4 Expected size of the generated/re-used data	A minimum of 100 subscribers are expected to be reached, according to D8.1. However, according to project implementation this number may vary.
1.5 Origin/provenance of the generated/re-used data	People voluntarily provide their contact details and agree to receive the project's email communications in their mailboxes. For that, they use the subscription form available on LEARN's website.
1.6 Data utility outside the project - To whom might your data be useful outside the project?	It will only be used internally, as a database of contacts to send email newsletters to.

2. Findability & Metadata

2.1 Data persistent identifier	N/A
<i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	
2.2 Metadata creation	Tags (customizable), email marketing (type of subscriber), source (of subscription), contact rating (automatically calculated), date added and last changed.
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	Metadata is generated in the MailChimp platform, using their standards.
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	YES. Keywords are available for the following metadata: tags (like "web subscriber"), email marketing (like "subscribed" or "cleaned") and source (like "API-Wordpress" or "Popup Form").
2.5 Possibility of harvesting/indexing the metadata	YES, metadata can be exported at any time in CSV format.

3. Accessibility

Repository	3.1 Data deposition in trusted repository	N/A
	3.2 Identifier assigned to data by the repository	N/A
Data	3.3 Data open access	N/A
	<i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	Open Access doesn't apply to this dataset, as this is not research data. This is personal data (contact details) that people only consented to be used for receiving LEARN's email communications. In compliance with GDPR, we are not allowed to publicly share this personal, sensitive data, nor to provide it to any third party.
	3.4 Data accessibility through a free and standardized access protocol	N/A
	3.5 Way to ascertain the identity of the person accessing the data	N/A

Metadata	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A
	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement "CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain." https://wiki.creativecommons.org/wiki/CC0	N/A
	3.9 Metadata availability on information that enables the user to access the data	N/A
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	After the project ends, data will be kept internally according to the European legislation on GDPR and the specific Horizon Europe requirements.
	3.11 Need of documentation/reference about any software to access or read the data	NO
4. Interoperability		
4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	MailChimp allows data transfer to/from WordPress and other marketing tools. Moreover, data can be exported in CVS, which is supported by many software tools.	
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	NO (data has very low complexity and depth, therefore this is not needed)	
4.5 Qualified references to other data included in the data (e.g., other data from your project, or datasets from previous research)	NO	
5. Re-use		
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	NO (data has very low complexity and depth, therefore this is not needed)	

5.2 Free availability of the data in the public domain to permit the widest re-use possible	NO. This is personal data (contact details) that people only consented to be used for receiving LEARN's email communications. In compliance with GDPR, we are not allowed to publicly share this personal, sensitive data, nor to re-use it for any purposes.
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	NO, data won't be licensed for re-use (see 5.2).
Usability of generated in the project by third parties, in particular after the end of the project	NO
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	In the subscription form, all fields will be mandatory (email address and first name). Moreover, all metadata will be configured to be automatically generated in the MailChimp platform.

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	The only costs are those associated with the personnel cost category.
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	They will be included in the person-months that FI has for WP8, according to the Grant Agreement.
6.3 How long they will be preserved after the end of the project	After the project ends, data will be kept for the required time according to the European legislation on GDPR and the specific Horizon Europe requirements.

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	MailChimp has its own measures for data security and counts on a Data Protection Officer to oversee compliance with European GDPR. It is also certified annually with the EU-US Privacy Shield Frameworks. Moreover, the exported data on FI Group's servers will be secured by the organisation's IT Department, which has a particular policy for
--	---

	data management, in compliance with ISO 27001, in which FI Group is certified.
7.2 Data storage in trusted repositories for long-term preservation and curation	NO. After the project ends, data will only be kept in FI Group's servers.
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	YES, the MailChimp platform will store data during the project's duration.
7.4 Procedure(s) in place to follow in the case of a data breach	The procedures of MailChimp and FI Group's IT Department will be used in case of data breach (see 7.1) and according to the GDPR regulation.
7.5 Data need encryption	Data might be encrypted if required by the procedures and certifications stated in 7.1.
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES. The European legislation on GDPR is complied. MailChimp follows the EU-US Privacy Shield Frameworks. In addition, FI Group is certified ISO-27001.
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	YES, FI Group's IT Department has specific procedures to avoid data loss, such as periodic backups.

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	This dataset contains personal data and will therefore comply with the European GDPR. Moreover, it will only be used for the purposes agreed by the subscribers in the subscription form.
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO. Subscribers only give consent to use data for receiving the project's email communications.
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	YES. The MailChimp platform and FI Group have their own procedures to protect the privacy, security and confidentiality of this personal data, as explained in 7.1.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not	YES. We won't ask for / collect any surnames, personal IDs or similar data, in order to avoid that individual persons can be identified. We will only collect an email address and a first name.

meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	After the project ends, data will be kept for the required time according to the European legislation on GDPR and the specific Horizon Europe requirements. They will be permanently eliminated afterwards.
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	The only information which will be shared with other partners will be the aggregated number of subscribers. Personal data of individual subscribers will not be shared.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Personal data will not be shared.

Performance analytics of online communication tools (website, social media and newsletters)

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Analytic data on the performance of the project's website, social media and newsletters, which shows their impact and targets reached (website visits, social media followers, newsletters opened...). It is numeric data. This is aggregated data that only includes aggregated amounts and never identifies any individuals. Data is stored in online databases of Google Analytics, Twitter, LinkedIn, YouTube and MailChimp. It is also stored offline in CVS/XLS files in FI Group's servers.

1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	This data will be used for T8.1 under WP8. In particular, to analyse the effectiveness and KPIs related to the website, social media and newsletters of the project, as stated in LEARN's Communication & Dissemination Plan (D8.1). This data will allow us to know the impact of the project's communication and the targets reached.
1.4 Expected size of the generated/re-used data	This data will take a size proportional to the number of communications that will be produced for the project's website, social media and newsletters, during the whole project.
1.5 Origin/provenance of the generated/re-used data	This aggregated data is calculated automatically and in real time by Google Analytics, Twitter, LinkedIn, YouTube and MailChimp, thanks to their tools to track the users' activity on their marketing platforms (such as cookies or tracked URLs). Data does not identify specific users, it is automatically anonymised and aggregated.
1.6 Data utility outside the project - To whom might your data be useful outside the project?	It will be used internally and shared with European commission members, as indicators to know the effectiveness of the project's communications and achievements of KPIs. In addition, it may be possible that some data will be included in project deliverables and therefore shared with general public.

2. Findability & Metadata

2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	Different metadata is automatically generated by these platforms. On the website, it can include: number of visits, number of users, duration of visits, country of users, source of visits... On social media, it can include: number of impressions or views, engagement rate,

	number of likes, number of followers, country of followers...
	On newsletters, it can include: number of email openings, number of clicks on links...
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	N/A
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO. This metadata is aggregated. Therefore, it is numeric. It includes qualifiable information, not quality information.
2.5 Possibility of harvesting/indexing the metadata	YES, metadata can be exported at any time in CSV/XLS formats.

3. Accessibility

Repository	3.1 Data deposition in trusted repository	N/A However, data is stored in an online database on Google Analytics, Twitter, LinkedIn, YouTube and MailChimp, standard platforms in the market for marketing purposes. Moreover, data will be exported periodically in CVS/XLS format and stored in FI Group's servers.
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	N/A Open Access doesn't apply to this dataset, as this is not research data. This is analytic data to know the performance of the project's online communication activities, which is meant to be used only internally in the project. Moreover, data that involves KPIs included in D8.1 will be reported to the European Commission.
	3.4 Data accessibility through a free and standardized access protocol	N/A
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A

Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	N/A
	3.9 Metadata availability on information that enables the user to access the data	N/A
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	After the project ends, data will be kept according to the specific Horizon Europe requirements.
	3.11 Need of documentation/reference about any software to access or read the data	NO

4. Interoperability

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	Google Analytics, Twitter, LinkedIn, YouTube and MailChimp are standard tools in the market for analytic data of websites, social media and email marketing. Therefore, their way to generate data is the standard in the market. They also support data transfer to other online marketing tools through API technology. Moreover, data can be exported in CVS/XLS, which is supported by many software tools.
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	NO, data follows standard formats and terminologies for digital marketing
4.5 Qualified references to other data included in the data (<i>e.g., other data from your project, or datasets from previous research</i>)	The online posts that are analysed, are often linked to the project's scope, activities and results. But the data itself does not include direct references to other datasets.

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	NO, data follows standard formats and terminologies for digital marketing
--	---

5.2 Free availability of the data in the public domain to permit the widest re-use possible	NO. This is analytic data to know the performance of the project's online communication activities, which is meant to be used only internally in the project.
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	NO, data won't be licensed for re-use (see 5.2).
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	All data is generated automatically by Google Analytics, Twitter, LinkedIn, YouTube and MailChimp, following their quality standards, that position them as leading platforms in digital marketing.

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	The only costs are those associated with the personnel cost category.
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	They will be included in the person-months that FI has for WP8, according to the Grant Agreement.
6.3 How long they will be preserved after the end of the project	After the project ends, data will be kept for the required time according to the Horizon Europe regulation, European legislation and according to each service providers protocols.

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Google Analytics, Twitter, LinkedIn, YouTube and MailChimp have their own measures for data security and data protection, as leading platforms for digital marketing. Moreover, the exported data on FI Group's servers will be secured by the organisation's IT Department, which has a particular policy for data management, in compliance with ISO 27001, in which FI Group is certified.
--	---

7.2 Data storage in trusted repositories for long-term preservation and curation	N/A However, after the project ends, data will be secured offline in FI Group's servers. It will also remain available online on Google Analytics, Twitter, LinkedIn, YouTube and MailChimp, leading platforms for digital marketing.
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	YES, Google Analytics, Twitter, LinkedIn, YouTube and MailChimp will generate and store this data.
7.4 Procedure(s) in place to follow in the case of a data breach	The procedures of Google Analytics, Twitter, LinkedIn, YouTube and MailChimp, as well as FI Group's IT Department, will be used in case of data breach (see 7.1).
7.5 Data need encryption	Data might be encrypted if required by the procedures and certifications stated in 7.1.
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department/</i>	YES. MailChimp follows the EU-US Privacy Shield Frameworks. Google Analytics, Twitter, LinkedIn and YouTube also follow international guidelines, which can be consulted on their respective websites. FI Group follows the ISO-27001 certification.
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software/</i>	YES, FI Group's IT Department has specific procedures to avoid data loss, such as periodic backups.

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	This dataset does not contain any personal or sensitive data. It is merely aggregated data on the performance of the project's online communications.
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO.
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	YES. Google Analytics, Twitter, LinkedIn, YouTube and MailChimp platforms, as well as FI Group, have their own procedures to protect privacy, security, and confidentiality, as explained in 7.1. However, there is no personal or sensitive data that needs special protection.

8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	YES. Data is aggregated, numeric and anonymised. There is no personal data, nor any data linked to any individuals.
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	N/A. There is no personal data.
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	N/A. There is no personal data.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	N/A

Contact details of workshop participants and final conference events

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Contact details of people who sign in to attend the project's workshops and/or final conference. These may include: name, surname, email address, phone number, organisation, role, dietary requirements, and consent to receive project communications, among others. Data will be stored on the project's internal Sharepoint and XLS files on FI Group's servers.

1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	This data will be used for T8.1 under WP8. In particular, for the project's workshop, final conference and any other event organised by the project, as stated in LEARN's Communication & Dissemination Plan (D8.1).
1.4 Expected size of the generated/re-used data	Reduced size expected. To be determined.
1.5 Origin/provenance of the generated/re-used data	People voluntarily provide their contact details in an online form where they sign up for the event. This form will be available on the project's website.
1.6 Data utility outside the project - To whom might your data be useful outside the project?	N/A It will be used internally, as a database of contacts to send the event communications.

2. Findability & Metadata

2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	Metadata may involve: name, surname, email address, phone number, organisation, role, dietary requirements, and consent to receive project communications, among others.
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	Metadata will be generated according to the fields that people fill in on the online form. They are described in section 2.2.
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	YES. Some fields on the sign-in form may require selecting pre-defined options, and will therefore generate keywords that will allow us to better analyse the resulting dataset.
2.5 Possibility of harvesting/indexing the metadata	YES, metadata will be stored in CVS/XLS formats.

3. Accessibility

 3.1 Data deposition in trusted repository	N/A
---	-----

	3.2 Identifier assigned to data by the repository	N/A
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	N/A Open Access doesn't apply to this dataset, as this is not research data. This is personal data (contact details) that people only consented to be used for receiving the event's communications. In compliance with GDPR, we are not allowed to publicly share this personal, sensitive data, nor to provide it to any third party.
	3.4 Data accessibility through a free and standardized access protocol	N/A
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	N/A
	3.9 Metadata availability on information that enables the user to access the data	N/A
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	After the project ends, data will be kept according to the European legislation on GDPR and the specific Horizon Europe requirements.
	3.11 Need of documentation/reference about any software to access or read the data	NO
4. Interoperability		
	4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable)</i>	Data is stored in CVS/XLS format, which are supported by many software tools.

to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)	
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	NO (data has very low complexity and depth, therefore this is not needed)
4.5 Qualified references to other data included in the data (e.g., other data from your project, or datasets from previous research)	NO

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	NO (data has very low complexity and depth, therefore this is not needed)
5.2 Free availability of the data in the public domain to permit the widest re-use possible	NO. This is personal data (contact details) that people only consented to be used for receiving the event's communications. In compliance with GDPR, we are not allowed to publicly share this personal, sensitive data, nor to re-use it for any purposes.
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	NO, data won't be licensed for re-use (see 5.2).
Usability of generated in the project by third parties, in particular after the end of the project	NO
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	In the sign-in form, all fields necessary for the event's management will be mandatory, in order to ensure that we collect the required data. Moreover, fields will have some formatting requirements to ensure that the resulting data adjusts to expectations: for example, the field for "email address" will only allow email addresses.

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	The only costs are those associated with the personnel cost category.
---	---

6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	They will be included in the person-months that FI has for WP8, according to the Grant Agreement.
6.3 How long they will be preserved after the end of the project	After the project ends, data will be kept for the required time according to the European legislation on GDPR and the specific Horizon Europe requirements.

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	The GDPR regulation is applied since project started. In addition, security protocols are implemented for each tool used for data storage: as a leading office tool for the business, institutional and educational environment, Microsoft Sharepoint (Office365) has its own, trusted procedures to ensure data security and protection. Moreover, data on FI Group's servers are secured by the organisation's IT Department, which has a particular policy for data management, in compliance with ISO 27001, in which FI Group is certified.
7.2 Data storage in trusted repositories for long-term preservation and curation	N/A
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	YES, Microsoft will be storage data collected.
7.4 Procedure(s) in place to follow in the case of a data breach	The procedures of Microsoft Sharepoint and FI Group's IT Department will be used in case of data breach (see 7.1) and complying with GDPR.
7.5 Data need encryption	Data might be encrypted if required by the procedures and certifications stated in 7.1.
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES. FI Group follows the ISO-27001 certification. Moreover, European legislation on GDPR will be complied. Microsoft Sharepoint also complies with the required international legislation and standards.
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	YES, FI Group's IT Department has specific procedures to avoid data loss, such as periodic backups. Microsoft Sharepoint also allows to recover previous versions or eliminated files.

8. Dataset ethical and legal impact	
8.1 Ethical or legal issues pertaining to the data	This dataset contains personal data and will therefore comply with the European GDPR. Moreover, it will only be used for the purposes stated in the sign-in form.
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO. People only give consent to use data for receiving the event's communications.
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	YES. Microsoft Sharepoint and FI Group have their own procedures to protect the privacy, security and confidentiality of this personal data, as explained in 7.1.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? <i>Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/</i>	We won't ask for / collect any personal ID numbers, in order to avoid that individual persons can be identified.
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	After the project ends, data will be kept for the required time according to the European legislation on GDPR and the specific Horizon Europe requirements. They will be permanently eliminated afterwards.
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? For example: encryption of personal data in transit, anonymization of personal data locally	Data might be shared with other partners in the project's Sharepoint, whose procedures for data security have already been mentioned. The purpose of this will be to establish connections with stakeholders.
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Data will not be allowed to be shared outside the project, with any third party.

Data related to the participatory policy making processes

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Personal data: contact details
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	WP8 /T8.2 Participatory policy making processes
1.4 Expected size of the generated/re-used data	Personal data: database of max. 50 contacts
1.5 Origin/provenance of the generated/re-used data	Contact details from personal networks or findable online
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Nobody

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>	N/A
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO
2.5 Possibility of harvesting/indexing the metadata	NO
3. Accessibility	
Repository	3.1 Data deposition in trusted repository
	NO

	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO The personal data will not be shared, restricted access
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	N/A
	3.9 Metadata availability on information that enables the user to access the data	N/A
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	N/A
	3.11 Need of documentation/reference about any software to access or read the data	N/A
4. Interoperability		
	4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	N/A

4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data (<i>e.g., other data from your project, or datasets from previous research</i>)	N/A
5. Re-use	
5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	N/A
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	N/A

6. Dataset financial management	
6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	N/A
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	N/A
6.3 How long they will be preserved after the end of the project	N/A

7. Dataset protection	
7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Storage on Sharepoint LEARN and VUB

7.2 Data storage in trusted repositories for long term preservation and curation	Until 5 years after the project ending
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	Inform the data subject and delete the personal data from the database
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES, VUB policies
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Sharepoint

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	YES
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	Only access by the partners involved in the task. Pseudonymisation of the answers of the participants in the final deliverable.
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	Yes
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which	5 years after the project ending Deletion of the file

it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	The partners Envirometrics, University of Arhus and University of Hasselt
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Yes

Research data related to the participatory policy making processes

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Research data
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	WP8 /T8.2 Participatory policy making processes
1.4 Expected size of the generated/re-used data	Research data: database of max. 150 contacts
1.5 Origin/provenance of the generated/re-used data	Online documents, document analysis, workshops
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Academia & policy makers

2. Findability & Metadata	
2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content." https://guides.lib.uw.edu/research/PID</i>	NO

2.2 Metadata creation		NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, <u>please outline what type of metadata will be created and how.</u>		
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use		NO
2.5 Possibility of harvesting/indexing the metadata		NO
3. Accessibility		
Repository	3.1 Data deposition in trusted repository	NO
	3.2 Identifier assigned to data by the repository	NO
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	NO
	3.4 Data accessibility through a free and standardized access protocol	NO
	3.5 Way to ascertain the identity of the person accessing the data	N/A
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement <i>"CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain."</i> https://wiki.creativecommons.org/wiki/CC0	N/A

3.9 Metadata availability on information that enables the user to access the data	N/A
3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	N/A
3.11 Need of documentation/reference about any software to access or read the data	N/A

4. Interoperability

4.1 How the data interoperability will be ensured <i>(What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)</i>	N/A
4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
4.5 Qualified references to other data included in the data <i>(e.g., other data from your project, or datasets from previous research)</i>	N/A

5. Re-use

5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided <i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	N/A
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	N/A

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	N/A
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	N/A
6.3 How long they will be preserved after the end of the project	N/A

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Storage on Sharepoint LEARN and VUB
7.2 Data storage in trusted repositories for long term preservation and curation	Until 5 years after the project ending
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	N/A
7.5 Data need encryption	NO
7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES, VUB policies
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Sharepoint

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?	N/A

For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))	
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	N/A (cfr. dataset 8.4)
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	5 years after the project ending Deletion of the file
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why? <i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	Transfer of research data among the partners Envirometrics, University of Arhus and University of Hasselt
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Yes

Subjective indoor air quality

1. Dataset summary	
1.1 Re-use of existing data and reason behind it	NO
1.2 Types and formats of generated/re-used data	Research data
1.3 Purpose of data generation/re-use in the project - Please state in which LEARN WP your data will be used/contribute to	WP2 /T2.4 Subjective indoor air quality
1.4 Expected size of the generated/re-used data	Research data: database of max. 300 research participants

1.5 Origin/provenance of the generated/re-used data	Surveys
1.6 Data utility outside the project - To whom might your data be useful outside the project?	Academia (& policy makers)

2. Findability & Metadata

2.1 Data persistent identifier <i>"A persistent identifier (PI or PID) is a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content."</i> https://guides.lib.uw.edu/research/PID	NO
2.2 Metadata creation	NO
2.3 Disciplinary/ general standards for the metadata - In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	
2.4 Availability of search keywords in the metadata - aiming to optimize the possibility of discovery and then potential re-use	NO
2.5 Possibility of harvesting/indexing the metadata	NO

3. Accessibility

Repository	3.1 Data deposition in trusted repository	I am not sure yet, it could be made open access if the data is fully anonymized
	3.2 Identifier assigned to data by the repository	Not sure yet, Zenodo might be an option
Data	3.3 Data open access <i>According to the EC, in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	Not know yet
	3.4 Data accessibility through a free and standardized access protocol	Not know yet

	3.5 Way to ascertain the identity of the person accessing the data	/
	3.6 Need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?	N/A
	3.7 Directory and file naming convention will be used	N/A
Metadata	3.8 Metadata open access and licence under a public domain dedication CC0, as per the Grant Agreement "CC0 is the "no copyright reserved" option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain." https://wiki.creativecommons.org/wiki/CC0	N/A
	3.9 Metadata availability on information that enables the user to access the data	N/A /NO
	3.10 Duration of data availability and findability, and availability of metadata after the data is no longer available.	N/A
	3.11 Need of documentation/reference about any software to access or read the data	N/A
4. Interoperability		
	4.1 How the data interoperability will be ensured (What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?)	N/A
	4.2 Provision of mappings to more commonly used ontologies in case the use of uncommon or the generation of project specific ontologies is unavoidable	N/A
	4.5 Qualified references to other data included in the data (e.g., other data from your project, or datasets from previous research)	N/A
5. Re-use		
	5.1 How documentation needed to validate data analysis and facilitate data re-use will be provided	N/A

<i>For example: readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.</i>	
5.2 Free availability of the data in the public domain to permit the widest re-use possible	N/A
5.3 Data licensing using standard reuse licenses, in line with the obligations set out in the Grant Agreement	N/A
Usability of generated in the project by third parties, in particular after the end of the project	N/A
5.4 How data quality assurance will be achieved – what process will you follow to ensure that?	N/A

6. Dataset financial management

6.1 Costs for making the data FAIR in your project and their potential value <i>For example: direct and indirect costs related to storage, archiving, re-use, security, etc.</i>	N/A
6.2 How they will be covered <i>Costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</i>	N/A
6.3 How long they will be preserved after the end of the project	N/A

7. Dataset protection

7.1 How data security will be achieved (<i>including data recovery as well as secure storage/archiving and transfer of sensitive data</i>)	Storage on Sharepoint LEARN and VUB
7.2 Data storage in trusted repositories for long term preservation and curation	Until 5 years after the project ending
7.3 Any third-party services/involvement or responsibility for data management or data storage activities	NO
7.4 Procedure(s) in place to follow in the case of a data breach	N/A
7.5 Data need encryption	NO

7.6 Any institutional data management guidelines, a data protection or security policy which you will follow <i>For example: Data protection officer/ Legal department/ Security department)</i>	YES, VUB policies
7.7 Any storage and backup strategy for the collected data <i>For example: Data is stored locally at the partner's premises and we use daily backup software)</i>	Sharepoint

8. Dataset ethical and legal impact

8.1 Ethical or legal issues pertaining to the data	NO
8.2 Inclusion of informed consent for data sharing and long-term preservation of personal data	NO
8.3 What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights? <i>For example: use of pseudonymized and anonymized datasets, access control in the storage service, encryption (method to be defined))</i>	N/A
8.4 Do you plan to use pseudonymisation and/or anonymization for any data category? Pseudonymisation of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. Fully 'anonymised' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on the processing of personal data under the General Data Protection Regulation (GDPR). Reference: https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation https://gdpr-info.eu/	Yes , pseudonymisation and anonymization by UH
8.5 The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. For how long will you store personal data and why? Please also state what protocols you will put in place to ensure you delete personal data that is no longer required to be stored?	5 years after the project ending Deletion of the file
8.6 Transfer and share of personal data – any protocols/procedures in place. With which partners are you planning to share personal data and why?	Transfer of research data among the partners Environmetrics, University of Arhus and University of Hasselt

<i>For example: encryption of personal data in transit, anonymization of personal data locally</i>	
8.7 For any personal data sharing and processing, will there be any central documentation to ensure protection of personal data and any agreement for sharing those data?	Yes