InChildHealth Deliverable

D1.1 Data Management Plan

Grant Agreement number	101056883
Action Acronym	InChildHealth
Action Title	IDENTIFYING DETERMINANTS FOR INDOOR AIR QUALITY AND THEIR HEALTH IMPACT IN ENVIRONMENTS FOR CHILDREN: MEASURES TO IMPROVE INDOOR AIR QUALITY AND REDUCE DISEASE BURDENS.
Type of action:	HORIZON RIA Grants
Version date of the Annex I against which the assessment will be made	20th May 2022
Start date of the project	1st September 2022
Due date of the deliverable	M6
Actual date of submission	28th of February 2023
Lead beneficiary for the deliverable	CSEM
Lead beneficiary for the Work Package	Aalto
Dissemination level of the deliverable	Public

Action coordinator's scientific representatives
PI, Prof. Heidi Salonen
heidi.salonen@aalto.fi
AALTO –KORKEAKOULUSÄÄTIÖ,
Aalto University School of Engineering
Department of Civil Engineering

Scientific Coordinator
Dr. Emmanuelle Castagnoli
emmanuelle.castagnoli@aalto.fi
AALTO –KORKEAKOULUSÄÄTIÖ,
Aalto University School of Engineering
Department of Civil Engineering

Funded by the European Union
UK Research and Innovation
Australian Government National Health and Medical Research Council NHMRC
Project funded by
Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

grants numbers: #101056883 (EU), #10040524 (UKRI), #22.00324 (SERI), #APP2017786 and #APP2008813 (NHMRC)
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, the SERI, the UKRI or the NHMRC. Neither the European Union nor the granting authorities can be held responsible for them.

Authors in alphabetical order		
Name	**Beneficiary**	**e-mail**
Marta Almeida	IST-ID	smarta@ctn.tecnico.ulisboa.pt
Rantala Aino-Kaisa	UOULU	aino-kaisa.rantala@oulu.fi
Stephan Dasen	CSEM	stephan.dasen@csem.ch
Judith Desmet	CSIC	jdeqsh@cid.idaea.es
Evangelia Diapouli	NCSR D	ldiapoul@ipta.demokritos.gr
Robert Ferguson	The University of Essex	rmwfer@essex.ac.uk
Ulla Haverinen-Shaughnessy	The University of Oulu	ulla.haverinen-shaughnessy@oulu.fi
Raquel Obregon	Sens Solutions	robregon@sens.solutions
Reza Safai-Naeeni	CSEM	reza.safai-naeeni@csem.ch
Carla Viegas	IPL	carla.viegas@estesl.ipl.pt

Abstract
The Data Management Plan outlines how the InChildHealth data will be handled both during research, and after the project. The goal of this deliverable is to consider the many aspects of data management, metadata generation, data preservation, and opening so that data are well-managed in the present and prepared for future preservation.

Contents

1. Introduction 4
1.1 Life cycle of this document 4
2. Data sources and collection 5
2.1 Aero-S3DP 5
2.2 Mini-Aero 6
2.3 AMR-S3DP 7
2.4 AirSensis 8
2.5 Naneos Partector 2 10
2.6 Dusttrak DRX 11
2.7 MA200 & AE51 12
2.8 Purple air sensor 12
2.9 CO2 sensor 13
2.10 Low volume sampler 14
2.11 InChildHealth AQ mobile sensing system 15
2.12 Integrated Risk Assessment Tool 16
2.13 Information and Alert System 17
2.14 LAMP 19
2.15 Portable HVAC and Filtration System 20
2.16 DELTA 21
2.17 SpO_2 Calculation Algorithm 23
2.18 Epidemiological Study 24
3. Data sharing and dissemination 28
3.1 Making data findable, including provisions for metadata 28
3.2 Making data openly accessible 29
3.3 Making data interoperable 29
3.4 Increase data re-use (through clarifying licences) 29
3.5 Dissemination Plan 30
4. Data storage, flow and backup 31
4.1 Storage 31
4.2 Flow 31
4.3 Backup 31
5. Data security and confidentiality 32
5.1 Data security 32

5.2 Data confidentiality 32
6. Media Tools 35
6.1 Project website 35
6.2 Social media 36
7. Ethical aspects and regulatory compliance 37
8. Conclusions 38

1. Introduction

InChildHealth includes practices for the open dissemination of research results and FAIR data, and citizen science. In line with the expected outcomes of this funding call, society will be impacted by providing scientifically based knowledge that will support the daily work and decision making of, for instance, authorities, cities, building owners and schools, resulting in healthier indoor environments. The network will publish all scientific articles, preferably with gold open access. In addition, all research outputs will be publicly available on the project's website, in institutional platforms, such as the Aalto platform and the European Commission's platform Open Research Europe.

Guidelines and training material according to our Dissemination and Communication Plan will be openly shared. A large amount of data will be produced, shared following the FAIR data principles[1] (Findable, Accessible, Interoperable, Reusable). The harmonization of sampling strategies, methods and techniques will guarantee high quality of the data and the comparability of the results, and protocols will be published. The curation, quality, interoperability, and reusability of the datasets will be implemented, and the citizen science approach will open the scientific process to the most important stakeholders of this research area, namely the children themselves, by exploring the involvement of youngsters and teachers in scientific research and the implications of these activities on and within society.

1.1 Life cycle of this document

The Data management Plan will be updated during the project development, depending on the data generated, preservation and dissemination foreseen. CSEM will organise meetings with the relevant partners every two months to ensure the timely update of information.

Role	Name	Affiliation	Email / Phone
DMP responsible	Stephan Dasen	CSEM	stephan.dasen@csem.ch

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

2. Data sources and collection

Inventories of instrumentation and methods will be established as an element in the harmonization of the data and comparability of the results.

Various types of new and unique qualitative (such as models, guidelines, recommendations, training material and policy briefs), quantitative (such as microbial, chemical, physical, physiological, and toxicological), and project administrative (such as meeting minutes, materials for reporting, cost sheets and legal documents) data will be collected and generated. Questionnaire surveys will produce data using web-based platforms. Measurement and online monitoring data will be produced mostly electronically as numbers and images. Final data files will be available as .csv, .xls, and .txt. Other data will consist of generally accepted formats (.docx and .pdf for documents/manuscripts, .png, .jpg, or .tif for images and illustrations, etc.). Audio will be saved in .aif or .wav format and videos in .avi or .mj2.

A substantial amount of sequence raw data results from paired-endform shotgun (metagenomes) and targeted marker gene sequencing will be produced. For example, in the Tier 0 analyses, 150 GB for 100 Million paired reads will be produced. Analysis of metagenomic data will use “metaWRAP”, and produce intermediate file formats such as .fasta, .sam, .bam, .GFF, and .txt. The final data will be a .fasta file for each metagenome-assembled genome (MAGs) with functional annotations in .GFF format. Abundances of taxa, functional genes and gene pathways across samples will be provided in .csv file format for analysis in R. For metabarcoding data, our processing pipeline will use a mixture “QIIME” and “VSEARCH”. File formats such as .gz, .fasta, .hist, .biom, .hdf5, .json, .txt, .tre, .html, .jpeg, .tsv will be produced. The final operational taxonomic unit (OTU) tables are provided as .csv and .biom files, with representative sequences in .fasta format. Downstream statistical analyses will be performed with R and SAS, file formats will be .csv, .rds, and .Rdata.

The following section will list all the sources of data that will be used in the context of the project. It may consist in devices that will acquire new data or the exploitation of currently available datasets.

2.1 Aero-S3DP

Responsible

Sens Solutions

Role

The Aero-S3DP will be customized for real-time monitoring of a broad range of gaseous and particulate pollutants such as CO, NO2, VOCs, PM, CO2, O3 and ~~Rn,~~ noise, T and RH.

Localization

Will be tested in schools.

Required inputs

No data input is required for this module.

Data Format & Size

Data format has yet to be decided.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that will be monitored with the Aero-S3DP sensing device.

Table 11. List of Parameters of Aero-S3DP monitoring device

Name	Acronym	Units	Range	Accuracy	Sampling Interval	Format	Size
Carbon Monoxide	CO	ppm	1 – 500	< ± 3 % of reading	20 s	tbd	tbd
Nitrogen Dioxide	NOx	ppm	1-20	< ± 2 %	0.5 s	tbd	tbd
Volatile organic compound	VOCs	ppm	1-1000	±15% m.v.	10 s	tbd	tbd
Particulate Matter	PM1,2.5,4,10	ug/m3	1-1000	PM1 & 2.5: +-10% PM4 & PM10: +-25%	1±0.04 s	tbd	tbd
Carbon Dioxide	CO2	ppm	400-10000	±(30 ppm + 3%)	20s	tbd	tbd
Ozone	O3	ppm	1-20	< ± 2 % of reading	30s	tbd	tbd
Radon	Rn	pCi/L	0.2 ~ 255 pCi/L (7~9,435 Bq/m³)	< ±10% (min. error <±0.5pCi/L)	10min	tbd	tbd
Noise		dB	100 Hz ~ 10 kHz	-45 to -39 dBV/Pa		tbd	tbd
Temperature	T	°C	-10 to 50	0.45°C	10s	tbd	tbd
Humidity	H	%RH	0-90	+-3%RH	20s	tbd	tbd

Data Utility

The data will be stored on cloud for further analysis.

2.2 Mini-Aero

Responsible

Sens Solutions

Role

A portable version of the mini-Aero will be developed, including a Bluetooth connection and a mobile app.

Localization

The system will be tested by students at school.

Required inputs

No data input is required for this module.

Data Format & Size

Data format has yet to be decided.

Data sensitivity

Are you acquiring/processing personal data? YES

- If yes, will the data be anonymized? YES

List of Parameters

The table below summarizes the set of parameters that will be monitored with the Mini-Aero sensing device.

Table 22. List of Parameters of Mini-Aero monitoring device

Name	Acronym	Units	Range	Accuracy	Sampling Interval	Format	Size
Nitrogen Dioxide	NOx	ppm	1-20	< ± 2 %	0.5 s	tbd	tbd
Particulate Matter	PM1, 2.5, 4,10	ug/m3	1-1000	PM1 & 2.5: +-10% PM4 & PM10: +-25%	1s	tbd	tbd
	TVOC	ppm	1-1000	±15% m.v.	10s	tbd	tbd
Temperature	T	℃	-10 to 50	0.45°C	10s	tbd	tbd
Humidity	H	%RH	0-90	+-3%RH	20s	tbd	tbd
GPS	GPS						

Data Utility

The data will be stored on cloud for further analysis.

2.3 AMR-S3DP

Responsible

Sens Solutions

Role

The AMR-S3DP will focus on detecting microorganisms using machine learning algorithms (Convolution Neural Networks - CNN) to identify specific VOC “fingerprints” from target organisms. For training and calibration, specific single organisms, and VOCs (five bacteria and five fungi) will be used in the aerosol lab of AIT. After the training phase, the system will be challenged with combinations of organisms. All systems will be evaluated by tests under reproducible lab conditions and in field studies, with specific compounds and by comparison to technologies of other providers.

Localization

The system will be tested in a laboratory.

Required inputs

No data input is required for this module.

Data Format & Size

Data format has yet to be decided.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that will be monitored with the AMR-S3DP sensing device.

Table 33. List of Parameters of AMR-S3DP monitoring device

Name	Acronym	Units	Range	Accuracy	Sampling Interval	Format	Size
Presence/ absence of target pathogen	P/A	NA	Presence absence	10 cfu	NA	tbd	tbd
Carbon Dioxide	CO2	ppm	400-10000	±(30 ppm + 3%)	20s	tbd	tbd
Volatile organic compound	VOC	ppm	1-1000	±15% m.v.		tbd	tbd
Particulate Matter	PM1,2.5,4,10	ug/m3	1-1000	PM1 & 2.5: +-10% PM4 & PM10: +-25%	1s	tbd	tbd
noise		dB	100 Hz ~ 10 kHz	-45 to -39 dBV/Pa		tbd	tbd
Temperature	T	°C	-10 to 50	0.45°C	10s	tbd	tbd
Humidity	H	%RH	0-90	+-3%RH	20s	tbd	tbd

Data Utility

The data will be stored on cloud for further analysis.

2.4 AirSensis

Responsible

NCSRD

Role

The AirSensis system provides real-time concentration data of PM10, PM2.5, CO, NO2 and CO2, and may be used for fixed and mobile applications since it will include a battery, WiFi/sim card connectivity and GPS.

Localization

The AirSensis system will be deployed during the T1 campaign in Athens (indoors and outdoors of 5 schools, 5 homes and 1 gym), as well as during the mobile measurements in Athens, Lisbon and/or Barcelona.

Required inputs

No inputs are required.

Data Format & Size

Data format and size are to be decided.

Data sensitivity

Are you acquiring/processing personal data? NO
Are you acquiring/processing personal data? YES

Yes. During fixed measurements (at schools and homes), the system will record the air pollutant concentrations in the indoor microenvironments, as well as the location (coordinates) of the measurement site. During the mobile application of the AirSensis system, the location of the study participant who will carry the device will be continuously monitored.

- If yes, will the data be anonymized? YES, all data collected will be anonymized.

List of Parameters

The table below summarizes the set of parameters that will be monitored with the AirSensis sensing device.

Table 44. List of Parameters of AirSensis monitoring device

Name	Acronym	Units	Range	Resolution	Sampling Interval	Format	Size
Particulate Matter (<10 µm)	PM10	µg/m3	0 – 6000 µg/m3	± 1 µg/m3	1 min	tbd	tbd
Particulate Matter (<2.5 µm)	PM2.5	µg/m3	0 – 6000 µg/m3	± 1 µg/m3	1 min	tbd	tbd
Carbon Monoxide	CO	ppm	1 – 1000 ppm	± 2 ppm or ± 5 %	1 min	tbd	tbd
Nitrogen Dioxide	NO2	ppm	0.1 - 10 ppm	± 0.1 ppm or ± 5 %	1 min	tbd	tbd
Carbon Dioxide	CO2	ppm	400 – 2000 ppm	± 50 ppm or ± 5 %	1 min	tbd	tbd
Temperature	T	° C	-20 - 70 °C	± 1 °C	1 min	tbd	tbd
Relative Humidity	RH	%	0 – 100 %	± 4 %	1 min	tbd	tbd
GPS	GPS						

Data Utility

This data will be used for the assessment of air quality in indoor ambient microenvironments (WP2) and for the quantification of personal exposure to air pollutants (WP4). The processed data will be also used for quantifying exposure to air pollutants in WP5.

2.5 Naneos Partector 2

Responsible

CSIC

Role

The Naneos Partector 2 is a stationary device that will be deployed for the measurement of particles. LDSA, number concentration, average particle diameter, total surface area and PM0.3 particle mass can be analyzed.

Localization

The Naneos Partector 2 will mainly be utilized for WP 2 data collection in Tier1 in schools and homes in Helsinki, Copenhagen, Barcelona, Athens, Vienna, Lisbon and Essex, during the years 2024 and2025.

Required inputs

No data input is required for this device.

Data Format & Size

Data is generated in .txt format and will be converted to csv files for time dependent analysis. Depending on the time interval used, the files will be a few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that can be monitored with the Naneos Partector 2.

Table 5 List of Parameters of the Naneos Partector 2

Name	Acronym	Units	Range	Accuracy	Sampling Interval
Lung deposited surface area	LDSA	$\mu m^2/cm$	Concentration: 0-12000 Particle size: 10 nm – 10 µm	± 30 %	1 s
Number concentration	N	pt/cm3	Concentration: 0-106 Particle size: 10-300 nm	± 30 %	1 s
Average particle diameter	d	nm	Particle size: 10-300 nm	± 30 %	1 s
Total surface area	S	µm2/cm3	Concentration: 0-50000 Particle size: 10-300 nm	± 30 %	1 s

Particle mass (<0.3 µm)	PM0.3	µg/m3	Concentration: 0-1000 Particle size: 10-300 nm	± 50 %	1 s

Data Utility

The data will be stored on cloud for further analysis.

2.6 DustTrak DRX

Responsible

CSIC

Role

The Dusttrak DRX is a stationary device that will be deployed for the measurement of particulate matter. PM1, PM2.5, PM10, respirable PM and total PM can be analyzed.

Localization

The Dusttrak DRX will mainly be utilized for WP 2 data collection in Tier1 in schools and homes in Helsinki, Copenhagen, Barcelona, Athens, Vienna, Lisbon and Essex, during the years 2024 and2025.

Required inputs

No data input is required for this device.

Data Format & Size

Data is generated in csv files. Depending on the time interval used, the files will be a few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that can be monitored with the Dusttrak DRX.

Table 6. List of Parameters of the Dusttrak DRX

Name	Acronym	Units	Range	Accuracy	Sampling Interval
Particulate matter	PM1, PM2.5, PM10, Resp, Total	mg/m3	0.001-150	± 0.1 % of reading or 0.001 mg/m3, whichever is greater	1 s

Data Utility

The data will be stored on cloud for further analysis.

2.7 MA200 & AE51

Responsible

CSIC

Role

The MA 200 and the AE51 are stationary devices that will be deployed for the measurement of black carbon concentration.

Localization

The MA 200 and the AE51 will mainly be utilized for WP 2 data collection in Tier1 in schools and homes in Helsinki, Copenhagen, Barcelona, Athens, Vienna, Lisbon and Essex, during the school year 2024-2025.

Required inputs

No data input is required for this device.

Data Format & Size

The MA200 generates a .csv data file and the AE51 generates .dat data files that can be converted to .xslx files using the software. The generated .csv/.xslx files can be used for time dependent analysis. Depending on the time interval used, the files will be a few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that can be monitored with the MA 200 and the AE51.

Table 7. List of Parameters of the MA 200 and the AE51

Name	Acronym	Units	Range	Accuracy	Sampling Interval
Black carbon concentration	BC	µg BC/m3	Particle size: 0-1 mg BC / m3	± 0.1 ug BC/m3	300 s

Data Utility

The data will be stored on cloud for further analysis.

2.8 Purple air sensor

Responsible

CSIC

Role

The purple air sensor is a stationary device that will be deployed for the measurement of particulate matter 2.5.

Localization

The purple air sensor will mainly be utilized for WP 2 data collection in Tier1 in schools and homes in Helsinki, Copenhagen, Barcelona, Athens, Vienna, Lisbon and Essex, during the years 2024 and2025.

Required inputs

No data input is required for this device.

Data Format & Size

Data is generated in .csv files. Depending on the time interval used, the files will be a few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that can be monitored with the purple air sensor.

Table 8. List of Parameters of the purple air sensor

Name	Acronym	Units	Range	Accuracy	Sampling Interval
Particulate matter 2.5	PM2.5	µg/m³	0-500 µg/m³	±10 µg/m³ at 0 to 100 µg/m³ and ±10 % at 100 to 500 µg/m³	1 s

Data Utility

The data will be stored on cloud for further analysis.

2.9 CO2 sensor

Responsible

CSIC

Role

The CO2 sensor is a stationary device that will be deployed for the measurement of carbon dioxide.

Localization

The CO2 sensor will mainly be utilized for WP 2 data collection in Tier1 in schools and homes in Helsinki, Copenhagen, Barcelona, Athens, Vienna, Lisbon and Essex, during the years 2024 and2025.

Required inputs

No data input is required for this device.

Data Format & Size

Data is generated in .csv files. Depending on the time interval used, the files will be a few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that can be monitored with the CO2 sensor.

Table 9. List of Parameters of the CO2 sensor

Name	Acronym	Units	Range	Accuracy	Sampling Interval
Carbon dioxide	CO_2	ppm	0.-5000 ppm	± 50 ppm or 5 % of reading, whichever is greater, from 3000 ppm: 7 % of reading	1 s

Data Utility

The data will be stored on cloud for further analysis.

2.10 Low volume sampler

Responsible

CSIC

Role

The low volume sampler is a stationary device that will be deployed for the collection of the vapour and the gas phase.

Localization

The low volume sampler will mainly be utilized for WP 2 data collection in Tier1 in schools and homes in Helsinki, Copenhagen, Barcelona, Athens, Vienna, Lisbon and Essex, during the years 2024 and 2025.

Required inputs

No data input is required for this device.

Data Format & Size

Data is generated in .xlsx for time dependent analysis. Depending on the time interval used, the files will be a few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that can be monitored with the low volume sampler.

Table 10. List of Parameters of the low volume sampler

Name	Acronym	Units	Range	Accuracy	Sampling Interval
plasticizers		ng/ m^3			
Flame retardants		ng/ m^3			

Biocides					
polycyclic aromatic hydrocarbon	**PAH**	pg/ m^3			

Data Utility

The data will be stored on cloud for further analysis.

2.11 InChildHealth AQ mobile sensing system

Responsible

IST-ID

Role

Wearable, lightweight, and silent sensor solutions are necessary to identify the individual exposure and determine the personalised dose to correlate to health effects. We will develop an air quality mobile exposure sensing system composed of a network of real-time sensor nodes to monitor children during their daily routine. This AQ sensing system takes advantage of concepts like big data analytics, IoT and citizen science and is a step forward into the exposome concept. Key features will be easy transport and wearability at all times, silent for use while working, studying and sleeping, equipped with GPS to identify the micro-environments, sensors for CO2, CO, CH2O, VOCs, PM2.5, PM10, T, RH and inclusion of a passive sampler for advanced analysis (wrist band).

Localization

The InchildHealth AQ sensing system will be used by the children during all their daily activities to assess individual exposure to air pollutants. The generated data will be stored in the cloud and presented to the users through the InChildHealth Information and Alert system.

Required inputs

The InChildHealth AQ mobile sensing system will use the sensor nodes **Mini-Aero** or **Air-Sensis** that will be installed in a backpack to be easily transported by the children.

Data Format & Size

The InchildHealth AQ mobile sensing system will produce georeferenced data for each measured pollutant with a time resolution of at least 1 minute.

Data sensitivity

Are you acquiring/processing personal data? YES

- If yes, will the data be anonymized? YES

List of Parameters

The list of parameters measured by the InChildHealth AQ sensing system are presented in Table 44 (related to Air-Sensis) and Table 22 (related to Mini-Aero).

Data Utility

The information generated by the InChildHealth AQ sensing system will be used by the project team to assess the individual exposure of children to air pollutants (WP2), to identify the main sources and daily activities affecting the exposure (WP2), to identify measures to reduce exposure (WP4) and to correlate with health effects (WP3). Moreover, the users will have access to the information, through the InChildHealth Information and Alert system, that will increase their awareness and help them to change daily routines and behaviours.

2.12 Integrated Risk Assessment Tool

Responsible

TUC

Role

The InChildHealth Integrated Risk Assessment tool will combine measurement results with historical data and multiple scientifically based models to evaluate the integrated exposure risk of children both at individual and collective (population) level in the microenvironments where the children are present.

Localization

The Tool will be available online and open to all end users.

Required inputs

The input data of the Tool will include ambient concentrations of pollutants, infiltration parameters for typical indoor microenvironments, emission factors for indoor sources, typical, city-specific, time-activity patterns for children, dosimetry parameters for the estimation of particles' deposited dose into the human organism and dose-response parameters, for the estimation of health impacts due to exposure to the different atmospheric pollutants.

Data Format & Size

To be defined

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The Tool will provide its outputs online and the end users may choose to inspect the results and/or save them for later use. In addition, the parameters that will be provided include a number of pollutants (some are listed in the table **Error! Reference source not found.**, the final list is yet to be decided) but also various types of outputs for each of these pollutants (e.g. indoor concentration, personal exposure, dose of PM deposited into the human organism, health impact parameters).

Table 11. List of Parameters of the integrated risk assessment tool

Name	Acronym	Units	Range	Resolution	Sampling Interval	Format	Size
Carbon dioxide	CO2	ppm	400-10000	±(30 ppm + 3%)	20s	tbd	tbd
Bacteria total		CFU/m2	tbd	tbd	1 day	tbd	tbd

Mass concentration of particles with diameters up to 10 µm	PM10	µg/m3	tbd	tbd	30 min	tbd	tbd
Mass concentration of particles with diameters up to 2.5 µm	PM2.5	µg/m3	tbd	tbd	30 min	tbd	tbd
Nitrogen dioxide	NO2	µg/m3	tbd	tbd	30 min	tbd	tbd
Carbon monoxide		µg/m3	tbd	tbd	30 min	tbd	tbd
Ozone	O3	µg/m3	tbd	tbd	30 min	tbd	tbd
Volatile organic compounds	VOC	µg/m3	tbd	tbd	tbd	tbd	tbd

Data Utility

The different sub-tasks of WP5 will provide data for sub-task 5.1, for the development of the InCHildHealth Tool. The Tool will be available online and provide data to end users on-line.

2.13 Information and Alert System

Responsible

IST-ID

Role

We will develop the InChildHealth information and alert system to engage children and the school community in improving IAQ. The system will present the data obtained by IAQ sensors to the educational community through a mobile application and informative screens. It is a bidirectional system where the educational community is informed about the status of the IAQ by the system, and at the same time, the school community feeds to the system information about their perceived IAQ, comfort and health status. This task will start with the conceptualization of the system, which will include: 1) the design of the operational platform that will consider the integration of data from any sensor technology; 2) the definition of permissions and functionalities for different types of users (e.g. students, teachers, building managers, project team); 3) the definition of the databases' management; 4) the development of IAQ indicators and index based on literature, WHO guidelines and the risk assessment results obtained in WP5; and 5) the development of a clear and user-friendly interface (important for target users and commercialization). Sens will develop the information and alert system and define protocols for actualization, maintenance, and QA/QC.

Localization

IST-ID and NCSRD will implement the system in a Portuguese and a Greek school to test, optimize and validate it and then deploy the system to other schools involved in the interventions. The data generated using the InChildHealth Information and Alert System will be stored in the cloud.

Required inputs

The InChildHealth project is developing four air quality monitoring instruments:

- The **Aero-S3DP** and **AMR-S3DP** will be installed in the pilot schools and will measure the concentrations of PM2.5, PM10, CO_2, NO_2, VOCs and, indirectly of microorganisms.
- The **Mini-Aero** and **Air-Sensis** will be part of the mobile air quality exposure system and will measure PM2.5, PM10, CO_2, CH_2O, CO, temperature, relative humidity and the user position through a GPS.

The data produced by the static and mobile instruments will be incorporated in the data platform that will provide the information to the project team. The ***InChildHealth user friendly information and alert system*** aims to make available the information to a broader audience - the school community.

Data Format & Size

A) Information given to the user through the information and alert system

A.1 Data produced by the **fixed monitoring sensors** will be presented to the users in the following formats:

- Bar chart with on-line concentrations of the pollutants measured in the pilot classrooms;
- Line chart with temporal patterns of the pollutants measured in the pilot classrooms;
- Radar charts with information about Key Performance Indicators (KPIs) that relate the measured parameters with the guidelines (also called threshold limit values, TLV). The system will present information for 4 KPIs: KPI-pollutants; KPI- IAQ; KPI-comfort and KPI-ventilation.
- Traffic lights that are alert systems that aim to inform the user when the concentrations exceed a TLV.

A.2 Information about the **characteristics of the classrooms** (area, number of occupants, etc) will be presented to the users as a table.

A.3 Data produced by the **mobile monitoring sensors** will be presented to the users in the following format:

- Map with the projection of the concentrations measured by the mobile sensors;
- Bar chart with average daily exposure levels.

B) Information provided by the user through the information and alert system

B.1 The user will fill in a questionnaire about perceived IAQ, comfort and health status. It will be a close ended questionnaire where the user will have to answer to questions using a scale from 1 to 4. The answers will be available only to the project team.

B.2 The users carrying the mobile sensors will use the system to fill and deliver their time activity diary

Data sensitivity

Are you acquiring/processing personal data? YES (GPS data)

- If yes, will the data be anonymized? YES

List of Parameters

Besides the IAQ data collected by the sensors developed within the project, and whose parameters were described elsewhere in this document, the InChildHealth information and alert system will also collect information from the users such as:

- perceived IAQ;
- perceived comfort;
- health status;
- Time activity patterns

Data Utility

The InchildHealth information and alert system was designed to be used by the entire school community: students, teachers, staff, school managers, and parents.

2.14 LAMP

Responsible

UEssex

Role

A highly efficient collection system with sample processing on-site (semi-automated sampling), combined with a detection system based on loop mediated isothermal amplification (LAMP), allows detection <60 minutes after the end of the sampling period. We will further improve the components of the system and validate them under stable conditions using a bioaerosol chamber. The classification as BSL-2 enables work directly with pathogens. The optimized components will be combined into a single molecular detection system to reduce operator.

Localization

Mostly tested in chamber studies, but it is possible some in-situ testing of the unit will take place at the test sites in WP2.

Required inputs

The device will use air samples, genetic material from target microorganisms.

Data Format

Data will be stored in .csv files.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that will be monitored with the LAMP sensing device.

Table 12. List of Parameters of LAMP monitoring device

Name	Acronym	Units	Range	Resolution	Sampling Interval	Format	Size
Presence/ absence of target pathogen	P/A	NA	Presence absence	1-10 gene target copies	NA	text/csv	<1 mb
Melt peak	MP	°C	NA	0.1	NA	text/csv	<1 mb

Data Utility

The data will be stored on cloud for further analysis.

2.15 Portable HVAC and Filtration System

Responsible

UOULU

Role

The effects of portable HVAC systems on reducing the spread of pollutants and, on a larger scale, controlled air distribution will be studied in a test room. The specifically developed portable HVAC systems with HEPA filters will be placed next to a potential infectious source (dummy) to prevent the spread of infectious diseases. The system will be connected to a real-time monitoring system (AMR-S3DP, Aero-S3DP), which can be programmed to regulate efficient HVAC activation and used as a potential preventive measure through a predictive model/machine learning algorithm. The application and scalability of this technique will be further tested in different settings (e.g. a school nurse clinic, classroom/teacher desk, and home(s)).

Localization

Data will be collected with the real-time monitoring system as described in 2.1 (Aero-S3DP) and/or 2.3 (AMR-S3DP).

Required inputs

Selection of systems to be tested.

Data Format & Size

Data format will be .csv or .txt files. The size of these data files is typically only few hundreds to thousands of kilobytes.

Data sensitivity

Are you acquiring/processing personal data? NO

List of Parameters

The table below summarizes the set of parameters that will be monitored by the HVAC and filtration system.

Table 13. List of Parameters of the HVAC and filtration system

Name	Acronym	Units	Range	Resolution	Sampling interval
Volatile organic compounds	VOCs	ppm	1-1000	±15% m.v.	10 s
Particulate Matter	PM1,2.5, 4,10	µg/m3	1-1000	PM1 & 2.5: +-10% PM4&PM10: +-25%	1±0.04 s
Carbon Dioxide	CO2	ppm	400-10000	±(30 ppm + 3%)	20s
Ozone	O3	ppm	1-20	< ± 2 % of reading	30s
Noise		dB	100Hz~10 kHz	-45 to -39 dBV/Pa	
Temperature	T	°C	-10 to 50	0.45°C	10s
Humidity	RH	%	0-100		10s

Data Utility

The data is related to WP4 (T4.2.1).

2.16 DELTA

Responsible

CSEM

Role

The DELTA is a wearable device developed by CSEM that offers real-time physiological data acquisition, enabling researchers to conduct in-depth analysis and visualization. It can be worn on the wrist or on the upper arm and allows to monitor data such as heart rate and heart rate variability, respiratory rate, blood oxygen saturation, activity monitoring, etc.

Localization

The DELTA devices will be primarily utilized in connection with Tier0 data collection in Helsinki, Finland, and intervention school sites running parallel, during the school year 2023-2024. The devices will be distributed among asthmatic students in the intervention and control schools. Later, the devices can be used in other cities as a part of WP5 (total exposure assessment) activities.

Required inputs

No data input is required for this module.

Data Format & Size

The transferred data format is .zip files containing a .bin file and its parsed .csv file.

The size of a file is roughly 10Mb per hours of recording.

Data sensitivity

Are you acquiring/processing personal data? YES

- If yes, will the data be anonymized? YES

List of Parameters

The table above summarizes the set of parameters that will be monitored with the DELTA sensing device.

Table 14. List of Parameters of the DELTA monitoring device

Name	Acronym	Units	Range	Resolution	Sampling Interval	Format	Size
Infrared raw PPG signal	PPG IR	-	$\pm 2^{20}$	1	25Hz	int32	4 Byte
Red raw PPG signal	PPG R	-	$\pm 2^{20}$	1	25Hz	int32	4 Byte
X-axis raw accelerometer signal	ACC X	mg	±8g	16mg	25Hz	int16	2 Byte

Y-axis raw accelerometer signal	ACC Y	mg	±8g	16mg	25Hz	int16	2 Byte
Z-axis raw accelerometer signal	ACC Z	mg	±8g	16mg	25Hz	int16	2 Byte
Heart rate	HR	BPM	0-210	1BPM	Configurable	float	4 Byte
Root mean square of successive NN intervals	RMSSD	ms	0-500	1ms	5 seconds	float	4 Byte
Standard deviation of the NN intervals	SDNN	ms	0-500	1ms	5 seconds	float	4 Byte
Percent of successive NN intervals that differs by more than 50ms	PNN50	%	0-100	1%	5 seconds	float	4 Byte
Percent of successive NN intervals that differs by more than 20ms	PNN20	%	0-100	1%	5 seconds	float	4 Byte
Low frequency power	LF	ms^2	0-250000	$1ms^2$	5 seconds	float	4 Byte
High frequency power	HF	ms^2	0-250000	$1ms^2$	5 seconds	float	4 Byte
Breathing Rate	BR	min^{-1}	0-60	$1min^{-1}$	5 seconds	float	4 Byte
Steps counter	steps	steps	$0\text{-}4*10^9$	1step	1 second	uint32	4 Byte
Walking distance	walking	m	$0\text{-}4*10^9$	1m	1 second	uint32	4 Byte
Running distance	running	m	$0\text{-}4*10^9$	1m	1 second	uint32	4 Byte
Energy expenditure	energy	calorie	$0\text{-}4*10^9$	1calorie	1 second	uint32	4 Byte
Speed	speed	km/h	0-30	0.1km/h	1 second	uint32	4 Byte
Activity class	class	N/A	N/A	N/A	1 second	uint8	1 Byte

Deep sleep	sleep	min	0-1440	1min	1 second	unit16	2 Byte

Data Utility

Raw signals (PPG & accelerometer signals) generated by this device will be used by the vital signs monitoring algorithm of CSEM that will calculate the SpO_2 (blood oxygenation). The processed data (all the other signals) will be stored on the cloud for further analysis. The data will be anonymised.

2.17 SpO_2 Calculation Algorithm

Responsible

CSEM

Role

This algorithm will process the raw signals from the DELTA device to calculate the SpO_2 of the asthmatic children.

Localization

The algorithm will run on the cloud. It has yet to be decided if it will run on CSEM premises or on a native cloud provider (e.g. AWS).

Required inputs

PPG and accelerometer signals from DELTA device.

Data Format & Size

The transferred data format is a .csv file.

The size of a file has yet to be estimated.

Data sensitivity

Are you acquiring/processing personal data? YES

- If yes, will the data be anonymized? YES

List of Parameters

The table below summarizes the set of parameters that will be monitored by the SpO2 Calculation Algorithm.

Table 15. List of Parameters of the vital signs monitoring algorithm

Name	Acronym	Units	Range	Resolution	Sampling Interval	Format	Size
Blood Oxygen Saturation	SpO_2	%	50-100	1%	1 second	float	4 Byte

Data Utility

The SpO_2 result will be stored on the cloud for further analysis. The data will be anonymised.

2.18 Epidemiological Study

Responsible

UOULU

Role

An epidemiological study conducted in schools in three European cities (Helsinki, Copenhagen and Barcelona) will assess the health impacts of indoor exposures in terms of respiratory infections (common cold, tonsillitis, sinusitis, otitis media, acute bronchitis and pneumonia), respiratory, eye and skin symptoms and asthma control, neurological and cognitional symptoms (working memory and attention trajectories, alertness, fatigue, headache, dizziness, and difficulty in concentration), and school absence related to these outcomes. The epidemiological study results will be complemented with interventions to assess the health effects under controlled settings, thus minimizing or eliminating potential confounding variables. Validated questionnaires inquiring data on personal and family characteristics, health information, daily activity, behavioural and dietary factors, details of the home environment will be used as well as the usual number of daily contacts with other people.

Localization

This questionnaire will be obtained from the schools of the three European cities (Helsinki, Copenhagen and Barcelona).

Required inputs

The questionnaire will be filled using REDCap (Research Electronic Data Capture) service by the parents. Cognitive tests will be conducted using web-based GORILLA platform. Data derived from using Mini-Aero and / or AirSensis will be uses for the epidemiological study. The data includes GPS data on some of the children.

Data Format & Size

Data can be downloaded as a .csv file.

Data sensitivity

Are you acquiring/processing personal data? YES

- If yes, will the data be anonymized? NO

List of Parameters

The tables below summarize the questions that will be asked to children's parents on a periodic base.

Table 16. Main outcomes obtained from baseline questionnaire on respiratory and allergic symptoms/diseases, and respiratory infections in children

Symptom/disease	Question
Asthma/Wheezing	
Wheezing (<12 months)	Has your child had wheezing or whistling in the chest in the last 12 months?
Doctor-diagnosed asthma	Has your child ever been diagnosed by a doctor as having asthma?
Medication for asthma or breathing difficulties	Did your child take any medicines for asthma or breathing difficulties (wheezing, chest tightness, shortness of breath) in the last 12 months?
Rhinitis	

Rhinitis (<12 months)	In the last 12 months, has your child had problems with sneezing, or a runny, or blocked nose when he/she DID NOT have a cold or the flu?
Rhinitis – with sneezing (<12 months)	If "yes": Please specify which of the symptoms your child had in the last 12 months, when he/she did not have cold or the flu? (Sneezing, Runny nose, Blocked nose)
Rhinitis – with runny nose (<12 months)	
Rhinitis – with blocked nose (<12 months)	
Doctor-diagnosed rhinitis	Has your child ever been diagnosed by a doctor as having allergic rhinitis (rhinitis due to the cat, dust) or seasonal allergic rhinitis (pollen rhinitis)?
Medication for nasal allergy/hay fever/ allergic rhinitis (<12 months)	Did your child take any medication against nasal allergy/hay fever/ allergic rhinitis in the last 12 months?
Eczema	
Dry skin (<12 months)	Has your child had dry skin in the last 12 months?
Itchy rash (<12 months)	Has your child had itchy rash at any time in the last 12 months?
Doctor-diagnosed eczema/atopic dermatitis	Has your child ever been diagnosed by a doctor as having eczema/atopic dermatitis?
Respiratory infections	
Cold/flu (<12 months)	Has your child had a common cold/flu in the last 12 months? How many times?
Otitis media (<12 months)	Has your child had otitis media in the last 12 months? How many times?
Tonsillitis/angina (<12 months)	Has your child had tonsillitis/angina in the last 12 months? How many times?
Maxillary sinusitis (<12 months)	Has your child had maxillary sinusitis in the last 12 months? How many times?
Bronchitis (<12 months)	Has your child had bronchitis in the last 12 months? How many times?
Pneumonia (<12 months)	Has your child had pneumonia in the last 12 months? How many times?

Table 17. Main outcomes obtained from the weekly questionnaire on respiratory health

Symptom/disease	Question
Respiratory infection during the week	Did your child have any respiratory infections during the past week?
Common cold (flu)	
Tonsillitis (angina)	
Maxillary sinusitis	
Otitis media (ear infection)	
Bronchitis	
Pneumonia	
Corona	
Influenza	
Respiratory symptoms	Did your child have any respiratory, nasal, eye, skin or some general symptoms during the past week?
Cough	
Secretion of phlegm	

Wheezing	
Shortness of breath	
Nasal symptoms (dry, itchy, stuffy or runny nose, sneezing)	
Eye symptoms	
Dry eyes	
Itchy eyes	
Eye irritation	
Watery eyes	
Bloodshot eyes	
Skin symptoms	
Dry or scaly skin	
Itchy skin	
Irritation or smarting of skin	
Redness of skin	
Blotchy red spots	
Soreness of pain of skin	
Numbness of skin	
Nettle rash or hives (swelling of the skin)	
Headache	
Tiredness	

Symptom/disease	Question
Asthma control	How is your asthma today?
	How much of a problem is your asthma when you run, exercise or play sports?
	Do you cough because of your asthma?
	Do you wake up during the night because of your asthma?
Asthma control (during the last 4 weeks)	During the last 4 weeks, how many days did your child have any daytime asthma symptoms?
	During the last 4 weeks, how many days did your child wheeze during the day because of asthma?
	During the last 4 weeks, how many days did your child wake up during the night because of asthma?

Table 19. Outcomes obtained from neurological/cognitional symptoms (self-assessment questionnaire with the cognitive test)

Neurological/Cognitional Outcomes	Question
Tiredness	During the last week, have you had any of the following symptoms?
Sleepiness	

Dizziness	
Headache	
Difficulty in concentrating	

Table 20. Outcomes obtained from the absenteeism assessment

Absenteeism assessment	Main Outcomes
Register of absenteeism from school	Number of absences per subject in one year.
	Number of absences per classroom in one year.
	Number of absences per school in one year.

Data Utility

Data from the questionnaires and cognitive tests will be used by the InChildHealth consortium.

3. Data sharing and dissemination

All partners will agree to the terms and conditions of the Material and Associated Data Transfer Agreement performed to be used by all the Consortium partners. In addition, all materials that were generated from differently funded Institutions or Consortia are subject to the terms and conditions of the InChildHealth Project.

Within the consortium, we will use a data sharing solution such as "SD connect" provided by the Finnish IT Center for Science CbSC, which is specific for sharing sensitive data in a secure way. The data with recognised long-term value will be archived with a storage period decided during the project, e.g., using the Finnish national Fairdata PAS service provided by CSC.

3.1 Making data findable, including provisions for metadata

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

The datasets of InChildHealth will be openly shared primarily on Zenodo which provides persistent identifiers (e.g., a DOI or URN) to promote data citation and generic standards for metadata. Monitoring data will also be published in the IPCHEM platform. Raw sequence data (e.g compressed .fastq files), genomes and metagenome assembled genomes (MAGs) will be uploaded to an appropriate database for these types of data (e.g. NCBI SRA, or ENA) and accession numbers provided in publications so that the data can easily be found, metadata associated within the records on the NCBI/ENA (e.g. sample location, data, sample type, the key words "air sample", and "bioaerosol" will be included) will make this data searchable at these websites.

The following naming convention applies to all files, programs, models, and data sets uploaded to the open data repository:

InChildHealth_101056883-<subject>_<place>_<detail>_<date>.<ext>

The prefix before "-" uniquely identifies the project. The other identifiers are:

- subject: sensor that generates the data. Examples: "PPG", "Temperature".
- place: identifies the place of data. Example: "School XX" .
- detail: optional information to complement subject and place.
- date: when the data is produced (year-month-day: YYYYMMDD-HH:XX).
- ext: file type extension. Directories are to be compressed in zip or tar.gz formats.

A version number may replace or complement the date if it makes sense, in addition to the version management provided by Zenodo or other tools.

Regarding sensitive personal data, such as health data, the European Code of Conduct for Research Integrity by ALLEA will be followed. The project Ethical Committee will ensure that the datasets are properly anonymised to permit their release. The data that cannot be efficiently anonymised will not be released, but metadata will be shared. Also, aggregated data (as a means of class, school, country or similar) can be transferred instead of personalized data to comply with ethics requirements. All shared datasets and metadata will be linked with persistent identifiers to institutional repositories and publicised on our website so that all are mapped in one place.

3.2 Making data openly accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

Open access strategies ensure the accessibility of data. Our IPR strategy will cover the IPR considerations on data accessibility which will be compatible with the protection of IPR, confidentiality obligations, ethical considerations, and the legitimate interests of the owner(s) of the results. The consortium reserves the possibility of registering intellectual property included in the results.

Some data may be made public at the end of the project depending on the agreement of test subjects and data anonymization (see section 5.2).

Raw sequence data will be made publicly available via the NCBI/ENA without restriction and with publication in a peer-reviewed journal, or within X years whichever is sooner, as is standard for this sort of data. Data will be in standard formats (e.g. compressed fastq files) so no specialist software will be required for its use, and appropriate metadata will be provided alongside it in accordance with FAIR principals. As this data will be available via both the NCBI and ENA without restriction it will not be possible to identify the person accessing the data.

During the Project and for a period of 1 year after the end of the Project, the dissemination of Results by one or several Parties including but not restricted to publications and presentations, shall be governed by the procedure of Article 17.4 of the Grant Agreement and its Annex 5, Section Dissemination, subject to the following provisions.

Prior notice of any planned publication shall be given to the other Parties at least 30 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement by written notice to the Coordinator and to the Party or Parties proposing the dissemination within 21 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted. To ensure the fulfilment of this requirement a shared file in Teams platform, dedicated to the scientific production in the scope of the project, will be available. The goal of this living document is to compile all the project outputs foreseen by all the partners from the consortium.

3.3 Making data interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

The InChildHealth guidelines for interoperability will include, e.g., data creation, storage, production, collection, distribution, and use. Datasets will be documented with descriptive metadata (e.g., the title/name of the data set(s), data creator(s)/author(s), abstract, keywords, date of data production, publisher, and persistent identifier). If needed, README files will be created to ascertain reusability, reading and interpretation of datasets. The metadata standards required by the repositories will be used. For instance, Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichments. Metadata will use a formal, accessible, shared, and broadly applicable language for knowledge representation. For instance, following Zenodo’s practices, JSON Schema can be used as an internal representation of metadata (which also offers export to other popular formats such as Dublin Core or MARCXML). References of metadata or data to other metadata or data will be qualified by a resolvable URL.

3.4 Increase data re-use (through clarifying licences)

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

Data will be richly described with a plurality of accurate and relevant attributes and access rights given through Creative Commons licences. The license recommended in InChildHealth for research data is Creative Commons Attribution 4.0 International (CC BY 4.0), and the waiver recommended for metadata is CC0 1.0 Universal Public Domain Dedication. The Integrated Risk Assessment tool will be available and openly accessible for data generation and validation of health impact mitigation strategies for other users.

3.5 Dissemination Plan

A dissemination plan is an essential component of any European project. It outlines the strategies and actions that will be taken to share the project results and outcomes with the relevant stakeholders, including the scientific community, policymakers, industry partners, and the general public.

Dissemination policies refer to a set of rules, regulations, and procedures that govern the sharing of research findings and outcomes with different stakeholders. The policies aim to ensure that research results are effectively communicated to the relevant audiences while also protecting the rights and interests of the researchers and their institutions. Some key elements of dissemination policies include:

- → **Open access**: Dissemination policies may require that research findings and outcomes are made freely available to the public through open-access publications or repositories.
- → **Intellectual property rights**: Dissemination policies may address issues related to intellectual property, such as copyright and patent protection, to ensure that researchers' rights are protected.
- → **Attribution and citation**: Dissemination policies may require that research results are properly attributed to the original authors and appropriately cited in subsequent publications.
- → **Data sharing**: Dissemination policies may require that research data are shared with other researchers to enable further analysis and replication of findings.
- → **Target audiences**: Dissemination policies may specify the target audiences for research results and the appropriate dissemination channels to ensure that the research findings are effectively communicated to the relevant stakeholders.
- → **Timing and frequency**: Dissemination policies may establish guidelines for the timing and frequency of dissemination activities to ensure that research findings are shared in a timely and appropriate manner.

In conclusion, dissemination policies are an important tool to ensure that research findings are effectively communicated to the relevant stakeholders while also protecting the rights and interests of the researchers and their institutions. By addressing issues related to open access, intellectual property, attribution, data sharing, target audiences, and timing and frequency, dissemination policies can help to maximize the impact of research outcomes and advance the state of knowledge in a given field.

All these important information's is well defined and described in InChildHealth Dissemination & Communication Plan.

4. Data storage, flow and backup

4.1 Storage

We will use data storage systems that consider the appropriate security level of the data in question. All consortium institutions hold high-security standards, and data is stored in password-protected environments that ensure secured access to the data only to authorized users involved in the research project. Within the consortium, we will use a data sharing solution such as Sensitive Data (SD) Connect provided by the Finnish IT Center for Science CSC, which is specific for sharing sensitive data in a secure way.

Data collected or acquired within the project will be stored in a secure IT environment behind a firewall at Aalto premises or in a secure cloud environment provided by Aalto's service providers. The data will be archived at the premises of Aalto for ten years after the project has finished.

Main centralized storage

Data management and data openness will align with FAIR Principles (Findable, Accessible, Interoperable and Re-usable). Such a strategy will promote joint use of data during the research and ensure the interpretation of data and their re-usability after the project. The data with recognised long-term value will be archived with a storage period decided during the project, e.g., using the Finnish national Fairdata PAS service provided by CSC. During the project research datasets will be available only to those project partners or project consortium members, who have been accredited by and their data usage has been approved by PI or authorized project consortium member.

IPCHEM

Genomic data will be stored at the University of Essex on two separate physical drives and one secure cloud location. The data will be shared across the consortium via the University of Essex high-performance computer cluster (CERES).

4.2 Flow

The field campaigns are organised in two tiers, namely Tier 0 and Tier 1.

Tier 0 is conducted in the three cities of Barcelona, Helsinki, and Copenhagen with a characterisation of the exposure of at least 2500 children per city. The targeted exposure assessment will be held in ten schools and ten classrooms per school for a campaign duration of one year.

Tier 1 will be performed in seven cities in Europe. T1 is covering five schools (with one indoor and one outdoor location per school), one sports hall and five homes in each of the seven cities. The measurements will be carried out for one week in each of the two seasons.

For both Tier 0 and Tier 1 data collection campaigns will respect the security and confidentiality considerations of section 5.

4.3 Backup

During the project, data will be stored primarily on secured institutional servers, which are supported by a snapshot feature and regular backups that enable automatic recoveries from unwanted deletions. Tape backups will provide system-level disaster recovery. We will use data storage systems that consider the appropriate security level of the data in question.

5. Data security and confidentiality

5.1 Data security

All consortium institutions hold high security standards, and data is stored in password-protected environments that ensure secured access to the data only to authorised users. Laptops include automatic data encryption, e.g., Bitlocker, and secure file transfer over the network with a VPN solution. Each institution's PI is responsible for controlling access to the project data on the institution's own servers. The consortium PI Heidi Salonen will control access to the project data in the shared SD connect solution.

5.2 Data confidentiality

Here is the list of the sensors who will collect personal data and will anonymize them. The goal is to detail here the processes to ensure anonymity of the data.

DELTA (CSEM)

The DELTA sensor will not save any information concerning a wearer's identity, but they are all uniquely identifiable. Therefore, it is essential that the sensors are handed to the wearers without traces of who received what device. To do so, they can simply be handed randomly.

If a registry is helping to know who received a sensor, in case of return management for example, one can have a list, even signed by the receiver, but it shall not contain the unique identifier of the sensor.

With such measure, we can consider the data collected by the DELTA sensor as truly anonymous, as no-one can know what data relates to whom.

SpO2 Calculation Algorithm (CSEM)

These algorithms will run on the DELTA sensor data collection, mentioned here above. Therefore, the anonymisation of the data described above will apply here as well.

Mini-Aero (Sens)

GPS data is the only sensitive data of this device. However, GPS data do not allow to identify one person. The devices will not contain the identity of the wearer. Moreover, they will be distributed randomly to the children, making impossible to trace the position to a dedicated child.

InChildHealth AQ mobile sensing system (IST-ID)

The entire data collected by the InChildHealth AQ mobile sensing system will be only available to the project team for research purposes. All research outputs (e.g. papers, presentations in conferences) will never present the results in an individualized format. Therefore, it will be impossible to identify children or trace their positions. The individual information will be made available to each user of the system for awareness and training purposes, which is essential to the proposed citizen science approach.

Information and Alert System (IST-ID)

The objective of the Information and Alert System is to make the data available to the schools and participants so they can change their habits in order to improve the Indoor Air Quality. We will define permissions and functionalities for different types of users (e.g. students, teachers, building managers, project team). Only the project team will have access to the entire database for research purposes. All research outputs (e.g. papers, presentations in conferences) will never present the results in an individualized format. The information from each school will only be available for the respective school community and after the authorization given by the school administration.

Epidemiological study

Epidemiological study obtains sensitive data using following measures:

1. Personal data, including child name, social security number, home address and parents' names and contact information, will be collected from the school records,
2. child's health data will be collected using REDCap service,
3. cognitive test will be conducted using web-based GORILLA platform and
4. ealth register information. In Finland this is maintained by the Finnish Institute of Health and Welfare and applied from the Finnish Social and Health Data Permit Authority Findata.

Questionnaires will be conducted using REDCap (Research Electonic Data Capture) service, where only researchers with an user account have access to. REDCap is a versatile and secure tool for building and managing online surveys and databases of research and suitable especially for handling of the sensitive research data. All the data captured in the REDCap is stored in the institutions' own servers where the data can be export for analysis. No project data is ever transmitted by REDCap from the institution to another institution or organization. The REDCap application employs various methods for data protection, such as filtering, sanitizing, data type checking and escaping of the submitted data.

GORILLA (www.gorilla.sc) is fully tooled experiment authoring and deployment platform, designed to resolve many timing issues and make reliable online experimentation open and accessible to a wider range of technical abilities. Currently, experiments have been conducted in Gorilla on a wide variety of topics, including assessing executive functioning and working memory among adults and children[2]. There is a separate "Code Editor" tab which provides a developmental environment to make experiments purely in code. The purpose of the "Code Editor" is to provide a secure and reliable service for hosting, data storage, and participant management for tasks written in code[3]. Data can be downloaded in csv format. Performance data has one trial per line for easy pivoting in Excel. Participant and performance data is downloaded separately in compliance with BPS guidelines.

Processing and storage of the primary data takes place in the each institution's own servers secured by firewalls, usernames and passwords. Only named members of the research team will have access to specified data. The data is processed coded in such a way that individual persons cannot be identified, personal data can no longer be attributed to a specific person without the use of additional information. Subject confidentiality is strictly held in trust by the participating investigators. Data that allows the identification of the subject will be kept separately from the remaining information and will be accessible only to the researcher who is responsible for the study. All the provided and collected data will be treated under a code to ensure information confidentiality. The study protocol, documentation, data, and all other information generated will be held in strict confidence. No information concerning the study, or the data will be released to any unauthorized third party.

The geocodes of children's home addresses, will be used to obtain, and characterize individual exposure estimates (air pollution, natural spaces and built-up areas, and distance to the nearest major road, motorway or highway). Only the environmental exposure estimates, and not the geocodes, will be linked to health information using a unique identification number.

[2] Anwyl-Irvine et al. 2020, Eschman et al. 2022
[3] Anwyl-Irvine et al. 2020

University of Oulu is the owner of the research register of the epidemiological study in Helsinki, ISGlobal of the epidemiological study in Barcelona and Aarhus University of the epidemiological study in Copenhagen. University of Oulu will allow research teams from Barcelona and Copenhagen to use the data in the Sensitive Data Desktop provided by the Finnish IT Center for Science CSC, which provides a secure workplace for collaborative research project. CSC's Sensitive Data Desktop will also be used for secondary use of health register data authorized by Findata.

6. Media Tools

The specifications of the developed InChildHealth website and social media are described in the InChildHealth Communication and Dissemination Plan.

6.1 Project website

The website consists of the following sub-pages:

- Home
- Partners
- Project overview
- Deliverables
- Publications&News
- Citizen of Science
- Contact

A screen shot of the InChildHealth website can be seen in Figure 11.

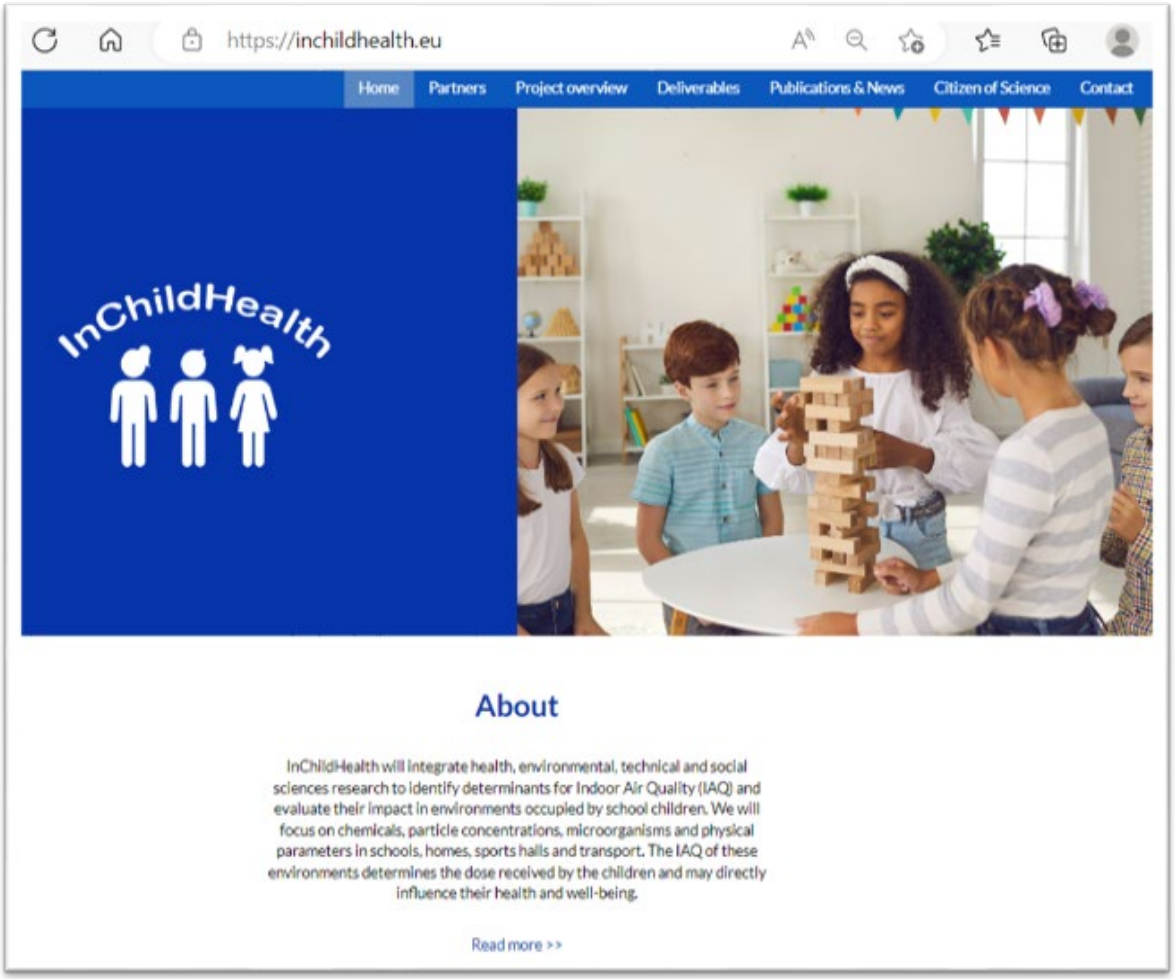

Figure 11. Screenshot of the InChildHealth Website

6.2 Social media

LinkedIn: a discussion group on LinkedIn has been set at the following url:

https://www.linkedin.com/company/inchildhealth/

The Privacy Policy is available at: https://privacy.linkedin.com/

Instagram: an official Instagram account will be created.

The Data Policy is available at: https://help.instagram.com/519522125107875/?helpref=uf_share

Twitter: An official twitter account will be created.

The Privacy Policy is available at: https://twitter.com/en/privacy

7. Ethical aspects and regulatory compliance

The research team will ensure that this study is conducted in full conformity with the principles set forth in the Declaration of Helsinki. The protocol will be submitted for approval request by the Ethical Commission. InChildHealth epidemiological study in Helsinki involves children aged 6-13 years recruited from ten primary schools. Research involves volunteer school children for studying the health parameters, including respiratory infections and cognition, with non-invasive techniques.

The participant's caregivers will provide informed consent through the school's information system before the start of the study, after receiving information document about the study in understandable language and terms. In addition, researchers will orally present the study for caregivers in a way that will be agreed with the schools. Caregivers will be contacted in Finnish and English and if necessary, also in other main languages in the schools, such as Russian, Somali, Estonian and Arabic. Participation in the study is voluntary, and participants can withdraw at any time without providing explanations. In the event of study's discontinuation, subjects will not be affected. They can ask questions and will receive understandable answers before taking any decision. Participants will be informed about their personal data protection rights and the risks, burdens, and benefits of their participation. Considering current information, participation in the study is not expected to cause any harm to participants and their wellbeing.

The InChildHealth consortium has appointed an Ethical Committee (EC), including Prof Jouni Jaakkola as a member. The Committee will conduct periodic ethical assessments of all project activities and prepare consortium guidelines for all ethical issues. They will identify the need for ethical permits and documentations and conduct an impact assessment for the use of personal data (Data protection impact assessment, DPIA). They will provide templates for documents required for research involving humans, such as Information sheets, Privacy Notice, or Informed Consent Forms. They will also specify the procedures for the treatment of personal data, which will be included in the DMP. All personal data collection, processing and storage will be carried out in accordance with the GDPR (EU 2016/679), national legislation, official guidelines, and good research ethics. Ethics committee of the hospital district implementing the study in each city (Pohde in Finland) will evaluate the research plan of the study in Helsinki.

8. Conclusions

This deliverable summarizes all the data providers of the InChildHealth project. It details all the data and how it will be used. It also outlies all the data flow exchanges. However, everything has not yet been set in stone in the project. For this reason, this document will be a living document, that will be periodically updated during the duration of the project.

This document also highlights the data that is critical (personalised data) and explains how it will be anonymised, or pseudonymised, to avoid privacy issues. Finally, more information is provided about how the data will be shared and published, and how we will follow the FAIR principles.