



IDEAL

INDOOR AIR QUALITY
HEALTH

DMP exchanged

Version 1.4

12.03.2024



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

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	3
Cluster details	3
IDEAL's projects.....	3
Cluster's aim	3
Document details	4
Abstract	4
Authors and contributors	4
Revision history	4
Disclaimer	5
2. Data Management and Analysis Working Group	6
2.1 Motivation and aims	6
2.3 Overview of types of data collected in the projects	6
3. Collaboration Topics	11
3.1 Approaches towards FAIR data	12
3.2 Public repositories with data that could be useful for the cluster aims	13
3.3 Protection and use of personal and sensitive data	14
3.4 Microbiome data.....	16
3.5 Data management best practices	17
3.6 Exposure times of in specific indoor environment.....	18
3.7 Harmonization and standardization of data collection techniques across EU labs	19
3.8 Definition of the hypothesis to be analysed with the data analysis	20
4. DMPs across different Projects	21
4.1 DMP common Timeline.....	21
4.2 DMP collaboration.....	22
5. Collaboration with IPCHEM.....	22
5.1 The role of IPCHEM	22
5.2 Objective and practical organisation	22
5.3 Identified actions	23
6. References.....	25

1. EXECUTIVE SUMMARY

Cluster details

Cluster title	Indoor Air Quality and Health
Cluster acronym	IDEAL
Dates and duration	September 1 st 2022 – August 31 st 2026 (48 months)
Horizon Europe call	HORIZON-HLTH-2021-ENVHLTH-02-02 Indoor air quality and health
Number of projects	7

IDEAL's projects

PROJECT	TITLE	GA NUMBER
K-HEALTHinAIR	Knowledge for improving indoor air quality and health	101057693
SynAir-G	Disrupting Noxious Synergies of Indoor Air Pollutants and their Impact on Childhood Health and Wellbeing	101057271
LEARN	Evaluation on indoor air quality for children around Europe	101057510
TwinAIR	Indoor air quality for healthy living	101057779
InChildHealth	Identifying determinants for indoor air quality and their health impact in environments for children: measures to improve indoor air quality and reduce disease burdens	101056883
Inquire	Identification of chemical and biological determinants, their sources, and strategies to promote healthier homes in Europe	101057499
EDIAQI	Evidence Driven Indoor Air Quality Improvement	101057497

Cluster's aim

In order to optimise synergies, avoid overlaps and increase the impact of the projects selected for funding under the call HORIZON-HLTH-2021-ENVHLTH-02-02 Indoor air quality and health, the projects have formed a cluster. Common cluster activities include the following:

1. Common cluster kick-off meeting of the six projects
2. Annual cluster meetings and periodic report of joint activities
3. Common dissemination and communication activities
4. Thematic workshops/trainings on issues of common interest
5. Working groups on topics of common interest
6. Advisory board



Document details

Deliverable Type	Report about actions of the IDEAL Cluster Working Group 2
Deliverable N°	-
Deliverable Title	DMP Exchanged
Name of Lead Partners for this deliverable	Jan Theunis (INQUIRE), Lukasz Wisniewski (TwinAIR)
Version	1.2
Contractual delivery date	31.08.2023
Actual delivery date	31.08.2023
Dissemination Level	Public
Purpose	This document was prepared in the framework of the work done in WG2 'Data analysis/management and protection'. It outlines the WG organisation providing information on common challenging topics between the cluster projects and the cluster's collaboration with IPCHEM.

Abstract

This deliverable provides the current status of the work being carried out within Ideal Cluster WG 2, which focuses on data management and analysis. It offers an overview of how the WG is organized, including information about sharing experiences to address common challenging topics that have been identified and prioritized for discussion so far, along with a description of the cooperation with IPCHEM.

Authors and contributors

Main authors: Lukasz Wisniewski (TwinAIR), Jan Theunis (INQUIRE)

Contributors: TwinAIR (Olympia Ageli, Antonis Papamanolis), INQUIRE (Valentina Marra, Anna Laura Iamiceli), K-HealthAIR (Cesar Mediavilla Martinez, Roberto Hernandez), LEARN (Evaggelia Samara, Vagelis Koufalis), SynAir-G (Maria Kampa), EDIAQI (Mario Lovrić, Eva Sophie Paulusberger, Kalle Kuusk), INCHILDHEALTH (Robert Ferguson)

Revision history

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

REVISION	DATE	DESCRIPTION	AUTHOR (PROJECT)
V.0.1	21.06.2023	First draft with the structure	Lukasz Wisniewski (TwinAIR)
V.0.2	06.07.2023	Revised structure and task distribution	Lukasz Wisniewski (TwinAIR)
V.0.3	21.08.2023	Integration of partners contributions	Lukasz Wisniewski (TwinAIR)



V.0.4	24.08.2023	Review of the semi final version	Jan Theunis (INQUIRE)
V.1.1	28.08.2023	Elaborating on comments and harmonising the document	Lukasz Wisniewski (TwinAIR)
V.1.2	29.08.2023	Final version before submission to the IDEAL coordinator	Jan Theunis (INQUIRE)
V.1.3	08.03.2024	Elaboration on comments from the PO	Lukasz Wisniewski (TwinAIR)
V.1.4	12.03.2024	Final check	Jan Theunis (INQUIRE)

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.



2. Data Management and Analysis Working Group

2.1 Motivation and aims

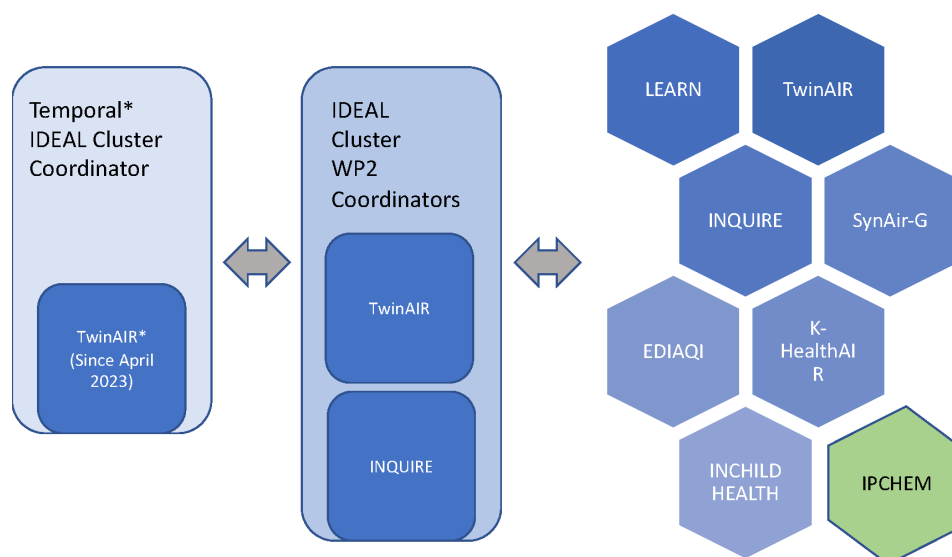
The primary objective of this working group 2 (WG2) is to establish a collaborative platform among 7 EU projects focused on the subject of indoor air quality. This platform aims to facilitate seamless knowledge exchange, idea sharing, and foster interaction among researchers specializing in data management and analysis. Considering the diverse scopes of the individual projects, ongoing collaboration and interaction will enable the identification of common topics and shared interests. The ultimate goal is to enhance the efficiency of research work and ensure that, in case of common interests, knowledge exchange and collaboration take place effortlessly.

Dealing with data encompasses a broad range of topics, including Sensor Technologies, Data Collection Methods, Data Quality Assurance, Data Sharing, Open Data, and many more. Therefore, the group aims to define and prioritize common topics relevant to the projects. These topics will be discussed within the platform, naturally stimulating interaction between project participants and providing an opportunity to receive feedback from different perspectives.

2.2 Collaboration Structure

The WG2 is being coordinated by the IDEAL Cluster coordinator, who is temporarily designated and takes care of all cluster activities. Since April 2023, it has been the TwinAIR Project. The work inside WG2 is coordinated by individuals permanently assigned to two projects, namely INQUIRE and TwinAIR. Both of these projects communicate with the cluster coordinator regarding progress status. The work inside WG2 is organized based on regular meetings with all seven projects and IPCHEM (the European Commission's Information Platform for Chemical Monitoring). The complete structure of the organization is depicted in the figure below.

Figure 1 WG2 Collaboration Structure



2.3 Overview of types of data collected in the projects

A data chart was elaborated by EDIAQI project to give an overview of the data collected among all seven IDEAL WG2 Cluster projects. This section provides an overview of the data collected in the second air quality data chart collection according to the following categories: project name, dataset

name, type of data collection, data types, chemical data, non-chemical data, and (project) partner. The data owned, provided, or (re)used by project partners in the chemical and non-chemical data section that could not be assigned to a specific consortium partner organization were dismissed in this iteration of data type collection, but will be specified in the third iteration.

The chart is equipped with a filter mechanism that improves the usability of the document. In the above mentioned further iteration the chart will be consolidated by a section on parameter ranges assigned to the chemical and non-chemical data provided. The collection and continuous adaption of data (categories) is an ongoing process to be redefined throughout the progress of WG2 Cluster project work. The specific timeline for this endeavour is defined more clearly in subsection 3.8.

Table 1 Air quality data collection from the whole WG2

Project Name	Dataset Name	Type of Data Collection	Venue of Data Collection	Data Types	Chemical Data	Partner
EDIAQI	Sensor data	Survey/interview/questionnaire	Residential/Homes/Households/Domestic	Sensorics Data	Air Pollutants	USEV
	Health data	Sensorics	Indoor Public Buildings/Non-Residential (schools, offices)	Building Type		TaITECH
	Biological data	Sampling	Outdoor Public Venues/Squares/Buildings/Places of Aggregation	Occupancy Utilization		LSA
	Environmental data	Database/Server		Frequency		RegionH
	Survey data			Geographic Location	VOCs	USEV
	Climate data			Validation Results		LSA
	Toxicological data			Public Sentiment on IAQ – standardized questionnaires	tVOCs	USEV
				Filter Content - sampled aerosols on filters		LSA
					Aromatic VOCs	USEV
						LSA
					Microplastics	T3.3 (UMOL)
						C3 Vilnius
					Radon	T3.1 (TROPOS)
					Gases	ASC
					Non-Chemical Data	
					Toxicity Data	IMROH
					Questionnaire Data	ANT
					Household Sample Data	ANT
					Biological Data	RegionH
						ANT
						TUG
					Biomonitoring Data	SCH
						IMROH
					Biobank Database Data	SCH

					Genetic Analysis Data	SCH
					Clinical Data	SCH
					Climate Data	TaITECH
						ASC
						USEV
						LSA
					Gravimetric Data	IMROH
K-HealthAIR	Sensor data	Survey/interview/questionnaire	Residential/Homes/Households/Domestic	Sensorics Data	Chemical Data	
	Health data	Sensorics	Indoor Public Buildings/Non-Residential (schools, offices)	Building Type	Air Pollutants	CIEMAT
	Biological data	Sampling	Outdoor Public Venues/Squares/Buildings/Places of Aggregation	Occupancy Utilization		Several partners
	Survey data			Frequency	VOCs	CIEMAT
				Geographic Location		NIO
					tvOCs	INBIOT
						M+H
					PAHs	WUT
					Radon	CIEMAT
						NIO
					Non-Chemical Data	
					Biological Data	CSIC
						WUT
LEARN	Sensor data	Survey/interview/questionnaire	Indoor Public Buildings/Non-Residential (schools, offices)	Filter Content - sampled aerosols on filters	Chemical Data	
	Health data	Sensorics	Outdoor Public Venues/Squares/Buildings/Places of Aggregation	Frequency	Air Pollutants	M & H
	Biological data	Purifiers		Geographic Location		KUL
	Toxicological data	Clinical tests		Public Sentiment on IAQ – standardized questionnaires		AU
	Survey data	Sampling		Sensorics Data	VOCs	KUL
		Models		Validation Results		M & H
				Laboratory results	Non-Chemical Data	
					Toxicity Data	INL
						AL
					Questionnaire Data	VUB
						AU
						ENV
						UH
					Biomonitoring Data	UH

						KUL
					Clinical Data	UH
InChildHealth	Sensor data	Survey/interview/questionnaire	Residential/Homes/Households/Domestic	Sensorics Data	Chemical Data	
	Health data	Sensorics	Indoor Public Buildings/Non-Residential (schools, offices)	Building Type	Air Pollutants	Uessex
	Biological data	Sampling	Outdoor Public Venues/Squares/Buildings/Places of Aggregation	Occupancy Utilization		AIT
	Survey data			Frequency		IPL
	Environmental data			Geographic Location		IST-ID
				Public Sentiment on IAQ – standardized questionnaires		NCSRSD
				Human interventions		Aalto
					VOCs	Uoulu
						AU
					tVOCs	CSIC
					Aromatic VOCs	Isglobal
					PAHs	TUC
INQUIRE	Sensor data	Sensorics	Residential/Homes/Households/Domestic	Sensorics Data	Chemical Data	
	Chemical data	Sampling		Non targeted screening	Air Pollutants	UCAM
	Toxicological data	Test chamber experiments		Perceived indoor air pollution	tVOCs	UCAM
	Biological data	Database/Server		Biological agents	VOCs	NILU
	Health data	Models		Geographic Location	SVOCs	NILU
	Geospatial data	Artificial mixtures		Building Type		SU
	Questionnaire data	Survey/interview/questionnaire		Occupancy Utilization	Phthaltes/PFRs/pesticides	UA
				Presence and use of energy systems, ventilation, appliances, consumer products, etc.	PAHs	ISS
				Demographics Data		VU
					Non-targeted screening (dust, products)	UMU
					Non-Chemical Data	
					Climate Data	UCAM
						UCAM
						MU
					Toxicity Data	KI
						VITO
						UH
					Biological data	UH
					Allergens	UH

TwinAIR	Survey data	Survey/interview/questionnaire	Indoor Public Buildings/Non-Residential (schools, offices)	Sensorics Data	Chemical Data	
	Sensor data	Sampling	Outdoor Public Venues/Squares/Buildings/Places of Aggregation	Filter Content - sampled aerosols on filters	Air pollutants	ISGlobal
	Biological data	Database/Server	Residential/Homes/Households/Domestic	Geographic Location		UCD
	Environmental data	Sensorics		Public Sentiment on IAQ – standardized questionnaires		IAAC
	Health data					UCAM
	Virological data					IRES
					VOCs	IAAC
					tVOCs	IAAC
						UCAM
						UCD
					Radon	IRES
						ISGlobal
					Non-Chemical Data	
					Biological Data	ISGlobal
					Viruses	UPAT
						CIIT
					Climate Data	IAAC
						UCAM
						UCD
					Noise levels	IAAC
SynAIR-G	Sensor data	Survey/interview/questionnaire	Residential/Homes/Households/Domestic	Sensorics Data	Chemical Data	
	Health data	Sensorics	Indoor Public Buildings/Non-Residential (schools, offices)	Frequency	Air Pollutants	CHUM
	Biological data	Sampling	Outdoor Public Venues/Squares/Buildings/Places of Aggregation	Geographic Location		FORTH
	Environmental data	Database/server		Public Sentiment on IAQ – standardized questionnaires		EPFL
	Survey data			Filter Content - sampled aerosols on filters	tVOCs	CHUM
	Climate data			Validation Results		FORTH
	Toxicological data					EPFL
					VOCs	CHUM
					Non-Chemical Data	
					Climate data	CHUM
						NAAVA
						NKUA
						UOULU



						CAIR
						UMAN
						UKER
					Questionnaire Data	NKUA
						CHUM
						UOULU
						CAIR
						UMAN
						UKER
					Biomonitoring Data	NKUA
						CHUM
						UOULU
						CAIR
						UMAN
						UKER
					Clinical Data	NKUA
						CHUM
						UOULU
						CAIR
						UMAN
						UKER
					Viruses	CNR

This air quality data chart shows the common terminology used and the umbrella terms defined to merge the data labels within a conjoined understanding. The terms highlighted in red have been brought up most recently as new categories to be discussed in a third iteration. As stated above the data (categories) for common terminology and collection are continuously discussed by project consortia and thus, redefined throughout the progress of WG2 Cluster project work. The specific timeline for this endeavour is defined more clearly in subsection 3.8.

3. Collaboration Topics

The section below contains topics identified so far within the WG2. All projects could propose topics for collaboration. These topics serve as the basis for further discussion and collaboration within WG2. This document contains the current state of discussion of the identified topics. Detailed objectives, outcomes and workplans need further elaboration and clarification. This will be done in the further course of 2023. A first effort towards prioritisation among these topics was done based on expressed interest of the projects in the proposed topics which is indicated under the heading “projects that expressed interest”. This prioritisation will be continued based on further discussion in the following regular WG meetings.



Each section is described according to the following pattern whenever possible:

Projects that expressed interest

Scope: delineate the topic, what does it include (and what not)?

Context: why is the topic important? is it relevant to address it in the cluster WG?

Objectives: what do you want to achieve in general?

Possible outcomes: what would exactly be achieved, produced, ...? which concrete results ("deliverables", "milestones")?

Actions: which concrete actions do you propose?

Timeline: when do you plan the actions? when will you achieve the outcomes?

3.1 Approaches towards FAIR data

Projects that expressed interest: SynAir-G, INQUIRE, LEARN, TwinAIR

Scope: delineate the topic, what does it include (and what not)?

FAIR principles are important for EU research projects to ensure findability, accessibility, interoperability and reusability of data.

However, the participants in the said projects find it difficult in several cases to understand and respect these principles, mainly due to data heterogeneity, technical challenges, unclear data governance policies and other constraints.

Despite these obstacles, the effort to make data FAIR is critical for realizing data's full potential for research, innovation, and decision-making across multiple domains. Overcoming these challenges necessitates stakeholder collaboration, the development of standardized data practices, and a commitment to transparency and openness in data management.

In this context, the scope of the activities includes the knowledge sharing and collaboration among the cluster projects on selected topics of interest to improve the knowledge and application of the FAIR principles.

Context: why is the topic important? is it relevant to address it in the cluster WG?

Adherence to the FAIR principles contributes to ensuring that data remains valuable and usable over time. In this regard, organizations and researchers can avoid data loss, obsolescence, and lack of documentation by making data findable, accessible, interoperable, and reusable. This is especially important as technology advances and new data analysis and integration methods emerge. The activities will be structured around the following pillars:

Objectives: what do you want to achieve in general?

Our goal is after the end of the activities, the participants to have gained an in-depth knowledge on interpretation and implementation of FAIR principles, such as the difference between FAIR and open data, interpretation and ways of implementation of FAIR criteria, importance to embed FAIR in all stages of a project and not only at the data reuse stage, levels of FAIRness.

Possible outcomes: The actions below will deliver a better understanding and alignment for the FAIR principles among participants. The workshop materials and the exchanges will ensure the delivery of FAIR data practice examples and will act as a reference for all cluster projects, as well as future initiatives.

Actions: Analyse and discuss the current state of methods and tools to implement FAIR principles

Workshop on the understanding of these principles and their implementation, demonstrating how the different principles and criteria can be applied to air quality and health data.

Knowledge and good practices sharing among projects.

The cooperation will take the form of presentations and discussions on the FAIR performance of different projects.

Timeline: The implementation timeline of the activities will run throughout the whole lifespan of the projects' duration.

3.2 Public repositories with data that could be useful for the cluster aims

Projects that expressed interest: LEARN, INQUIRE, K-HealthAir, EDIAQI, SynAir-G

Scope: Public repositories with data relevant to the cluster's objectives of studying indoor air pollution and its impact on human health encompass a broad range of resources. These repositories may include datasets on air quality measurements, pollutant concentrations, health outcomes, epidemiological studies, and exposure assessments. They serve as centralized platforms for researchers, policymakers, and stakeholders to access, share, and analyse data to advance knowledge in the field. While these repositories provide a wealth of information, it's important to note that they may not cover proprietary or sensitive data, and researchers may need to comply with specific data usage guidelines and protocols when accessing and utilizing the data.

Context: The topic of public repositories with data relevant to indoor air pollution and its impact on human health is of utmost importance within the cluster working group focused on this issue. Indoor and outdoor air pollution poses significant health risks, affecting individuals worldwide, particularly in densely populated areas where people spend a significant portion of their time indoors. Understanding the sources, pollutants, exposure levels, and associated health outcomes is crucial for developing effective mitigation strategies, policies, and interventions to safeguard human health. Public repositories serve as vital resources for researchers and stakeholders to access and share data, facilitating collaborative efforts, evidence-based decision-making, and the development of targeted interventions. By addressing this topic within the cluster working group, researchers can leverage the valuable data available in public repositories to enhance their understanding of indoor air pollution, support informed decision-making, and drive meaningful advancements in public health policies and practices.

Objectives: The general objectives of addressing the topic of public repositories with relevant data in the cluster working group are as follows:

1. **Enhance Knowledge:** By utilizing data from public repositories, the aim is to deepen the understanding of indoor air pollution and its effects on human health. This includes identifying key sources, pollutants, exposure patterns, and associated health outcomes.
2. **Collaborative Insights:** Foster collaboration among researchers, policymakers, and stakeholders by sharing and analysing data from public repositories. This collaborative approach can lead to new insights, cross-validation of findings, and a more comprehensive understanding of indoor air pollution.
3. **Evidence-based Decision-making:** Utilize the data available in public repositories to inform evidence-based decision-making. By analysing the data, researchers and policymakers can develop effective strategies, policies, and interventions to mitigate the adverse effects of indoor air pollution on human health.
4. **Inform Policies and Interventions:** Translate the findings derived from public repositories' data into actionable recommendations for policymakers and stakeholders. The objective is to contribute to the development of targeted interventions, regulations, and policies that aim to improve indoor air quality and protect human health.

Overall, the objectives are to leverage the wealth of data available in public repositories to enhance knowledge, drive collaboration, inform evidence-based decision-making, and ultimately improve the

understanding and management of indoor air pollution and its impact on human health within the cluster working group.

More specifically, the WG will have the primary objective of identifying pertinent and easily accessible repositories containing air pollution data that align with the goals of the projects. The WG will then proceed to update these repositories with the data generated from the projects. Additionally, the WG will engage in collaborative efforts with the IPCHEM team to devise appropriate formats for storing and sharing the data via the IPCHEM Portal. This process will draw inspiration from the practices of analogous repositories that have proven successful in similar contexts.

Possible outcomes: By addressing the subject of public repositories containing pertinent data concerning indoor air pollution and its repercussions on human health within the cluster WG, we can comprehensively review all available data encompassing indoor and outdoor air quality. This entails examining sources of indoor air pollution, levels of pollutant concentrations, health outcomes, and exposure assessments. Through this, our aim is to identify correlations between these factors and health indicators. Furthermore, we intend to conduct comparative analyses of outcomes following the implementation of specific policies aimed at enhancing indoor and outdoor air quality.

As an illustrative instance, consider the LEARN project, which is dedicated to investigating the influence of indoor air quality on the cognitive function and general health of children aged 9-12. By harnessing the potential of public repositories, researchers have the opportunity to establish connections between their own findings and the outcomes derived from other projects that focus on diverse age groups. This collaborative approach enhances the overall understanding of how indoor air quality impacts health across different demographic segments.

Ultimately, the possible outcomes of leveraging public repositories will be the generation of robust scientific evidence, informed policymaking, and the implementation of effective measures to address indoor and outdoor air pollution and safeguard human health.

Actions: ENV collected databases from various organizations and research centres throughout Europe, including the European Commission, agencies, and member state bodies. These databases were compiled in an Excel file, along with the dataset name, origin, website, and a brief description. The file was then shared on the cluster SharePoint. All cluster partners were requested to update the Excel file with relevant databases, publications, or project-generated publications. The purpose of this effort was to encourage data sharing and foster collaborations among the projects.

Timeline: Due to most projects being in their initial or second year, a significant amount of data has not yet been collected or generated. Consequently, this task is projected to conclude at the end of the projects. It serves as a supportive measure that will assist researchers throughout the entire project timeline as the database repository will be consistently updated with new databases as references.

3.3 Protection and use of personal and sensitive data

Projects that expressed interest: INQUIRE, LEARN, SynAir-G

Scope: This topic deals with the efficient and effective protection of various data categories according to the European Union's General Data Protection Regulation (GDPR) regulation, i.e., personal data and sensitive data. Several IDEAL projects collect personal data, such as data on individuals' homes, behavior, or lifestyle patterns, including the special category of personal data which is health-related data (according to Article 9 of the GDPR).

Context: The topic was selected as it was considered an important aspect of the projects' implementation, ensuring the following:

- **Privacy Protection:** To protect individuals' privacy, project partners should implement appropriate measures for sensitive and personal data, which usually contains private and

confidential information. This could include implementing techniques such as encryption, access controls, and data minimisation.

- **Data Security:** Sensitive and personal data can pose potential harm, discrimination, or significant consequences if accessed or misused. Organizations could enforce stronger security measures, such as robust authentication, network security, and data encryption, to prevent unauthorized access and data breaches by discerning between these data types.
- **Legal and Regulatory Compliance:** The GDPR outlines distinct categories of personal data and imposes specific obligations on organizations handling such data. Understanding the difference between sensitive and personal data is crucial for compliance with these regulations and building trust among research participants.
- **Risk Assessment:** By identifying sensitive data, organizations can also focus on implementing enhanced security measures and risk mitigation strategies to reduce the impact of unauthorized access or data breaches.
- **Informed Consent and User Control:** Differentiating between sensitive and personal data empowers individuals to make informed decisions about data sharing. This understanding will allow project partners to provide clear explanations and obtain explicit consent when collecting and processing sensitive data.

Objectives: Understanding the difference between sensitive and personal data, and the practical implications for data protection, results in better data protection, increased compliance, effective risk management, user trust, reduced data collection, and robust data governance. These results will contribute to a more secure and privacy-conscious data handling approach, which benefits both individuals and organizations.

Possible outcomes: The actions below will deliver a better understanding for implementing targeted and efficient data protection measures for personal and sensitive data. The workshops material will act as a reference for all cluster projects.

Actions: In order to facilitate effective collaboration, we have identified the following important aspects:

- Analysis of GDPR requirements and regulations in terms of the differences between personal and sensitive data, and the practical consequences for data protection.
- Tools and Processes in place to ensure the abovementioned data protection.
- Identification of any implementation gaps or areas that require further attention.

The proposed collaboration format will be based on sharing project updates and findings through informative presentations on specific topics. This format will encourage knowledge sharing, discussion, and feedback among all stakeholders.

Additionally, we plan to hold two workshops during the WG2 timeline. The workshop topics will be related to the difference of the personal and sensitive data as well as on compliance checklists. These workshops will provide opportunities for in-depth discussions, brainstorming, and collaboration on critical aspects of the said topic. They will serve as platforms to address any challenges, refine strategies, and align our efforts towards successful outcomes.

We believe that this collaborative approach, coupled with presentations and workshops, will foster effective communication, knowledge exchange, and collective decision-making throughout the project.

Timeline: Since these topics are important during the first period of the projects' implementation, the plan is to organize the two workshops until Q2 2024.

The related workshops material will be made available right after the abovementioned dates.

3.4 Microbiome data

Projects that expressed interest: LEARN, INCHILDHEALTH, INQUIRE, EDIAQI

Scope: The proliferation of high-throughput sequencing methods has led to a wealth of data on microbiomes across a range of environments, including in indoor air, and how these influence our health. Much of these data are publicly available in data repositories (see section 3.3), however, there are no standard methods or guidelines on how to collect and analyse these data. If these data could be combined they could vastly improve our understanding of the interaction between the microbiome and human health. The scope of this subgroup is to explore best practice for comparing and combining microbiome data where it is collected by different research groups (as in this cluster), and establish best practice for sample collection, processing, and analysis that will support both more coordination of research, and in integration of these data types into regulation and public health policy.

Context: Existing standards and guidelines for characterising bioaerosols and microbiome data in indoor air rely on short 'snap-shot' sampling, often using culture and microscopy for fungal and bacterial bioaerosols as they are simple and relatively low-cost. But there are several inherent limitations with culture-based methods, such as limited data, an underestimation or misrepresentation of the community's present making it hard to determine the health effects. Molecular tools, (e.g. High Throughput Sequencing), have significantly advanced bioaerosol research, enabling more robust associations between bioaerosols and health to be made. Yet, these bioaerosol/microbiome sampling and analysis methods are being used with no clear guidelines and recommendations. By addressing this topic within the cluster working group we will be able to enhance comparability across existing microbiome datasets and set guidelines and best practice for the generation and use of microbiome data by researchers and regulators. Moreover, this will support the evidence base required for officials to set guidelines and thresholds for improved bioaerosol biomonitoring that will underpin meaningful advancements in public health policies and practices.

Objectives:

1. Review paper on the methods used for high-throughput sequencing in indoor environments/air.
2. Develop recommendations for best practice in microbiome research to ensure comparability of data.
3. Engage with policy makers and other stakeholders (for example by preparation of a white paper on how microbiome data can be leveraged to support public health policy and regulation.

Possible outcomes: Improved comparability of microbiome data supporting improved collaboration and ability to combine existing and future datasets to make more robust inference on the interaction between microbiome/indoor air and human health.

Actions: A systematic review will be conducted, and the data synthesised and published in a peer-reviewed journal. This will be followed by/ or include collaboration on a set of recommendations for best practice. Our findings will then be published in a white paper, we will also take steps to engage with policy makers in other ways, at the national as well as at the European level.

Timeline: A concrete timeline for the mentioned actions still needs to be defined by the partners interested in this topic.



3.5 Data management best practices

Projects that expressed interest: SynAIR-G, LEARN, TwinAIR, INQUIRE

Scope: Data management is a critical aspect of modern organizations, as data has become a valuable strategic asset. Effective data management practices are essential for maintaining data integrity, ensuring data quality, and harnessing the power of data for informed decision-making.

As technology advancements continue to generate vast amounts of data, organizations must effectively manage and utilize this data to gain a competitive edge, drive informed decision-making, and achieve their business objectives. Data management best practices provide organizations with a framework and guidelines to handle data efficiently, ensuring its quality, integrity, security, and availability throughout its lifecycle.

Context: Data management is important for several reasons:

- **Maximizing Data Value:** Data management enables organizations to extract valuable insights from their data, leading to improved decision-making, operational efficiency, and identifying new business opportunities.
- **Mitigating Risks:** Implementing data management best practices helps organizations protect sensitive information, mitigate data breaches, address security vulnerabilities, and comply with regulations, reducing the risk of legal consequences and reputational damage.
- **Enhancing Customer Experience:** By collecting and analyzing customer data, organizations can personalize their offerings, improve customer support, and increase customer satisfaction and loyalty.
- **Regulatory Compliance:** Data management practices ensure organizations comply with data protection and privacy regulations, maintaining customer trust and avoiding legal penalties.
- **Efficient Operations and Cost Savings:** Effective data management streamlines processes, reduces errors, optimizes resource utilization, and helps organizations optimize storage costs and adhere to data retention policies.
- **Decision-Making and Strategy Formulation:** Accurate and accessible data enables stakeholders to make informed decisions, identify trends, forecast market changes, and develop effective strategies.

Objectives: Implementation of best practices to help organizations optimize their data handling processes, mitigate risks, and unlock the full potential of their data assets. Here are some examples of best practices:

- **Data Governance:** Data governance sets the foundation for data management best practices. It encompasses establishing policies, procedures, and guidelines to ensure data availability, usability, integrity, and security. Key topics within data governance include:
- **Data Ownership and Accountability:** Clearly defining roles and responsibilities for data ownership.
- **Data Classification:** Categorizing data based on sensitivity and regulatory requirements.
- **Data Lifecycle Management:** Managing data from creation to disposal, including storage, analysis, and retirement.
- **Data Security:** Data security is crucial for protecting sensitive information from unauthorized access, breaches, or theft. It involves implementing robust security measures and adhering to privacy regulations. The scope of data security best practices includes:
- **Access Controls:** Implementing role-based access controls and user authentication.
- **Encryption:** Protecting data at rest and in transit through encryption.
- **Security Audits:** Conducting regular audits to ensure compliance and identify vulnerabilities.
- **Data Storage:** Efficient data storage practices ensure data availability, scalability, and cost-effectiveness. Key topics within data storage best practices include:
- **Storage Technologies:** Selecting appropriate storage solutions based on scalability, accessibility, performance, and cost.

- **Data Backup and Recovery:** Implementing processes for regular data backups and disaster recovery.
- **Data Archiving:** Securely storing data no longer actively used but needed for compliance or historical purposes.
- **Data Integration:** Data integration is essential for creating a unified view of data and enabling seamless data flow across systems. The scope of data integration best practices includes:
 - **Data Integration Processes:** Establishing standardized processes for data mapping, transformation, and validation.
 - **Data Quality Management:** Ensuring data consistency and quality through cleansing and validation.
 - **Data Integration Tools:** Leveraging automation tools to streamline data integration and reduce manual efforts.
- **Data Analysis:** Effective data analysis practices enable organizations to extract valuable insights from their data. Key topics within data analysis best practices include:
 - **Data Preparation and Cleansing:** Cleaning, transforming, and aggregating data for accurate analysis.
 - **Advanced Analytics Techniques:** Using machine learning, data mining, and predictive modeling for valuable insights.
- **Data Visualization:** Presenting data in visually appealing ways to aid understanding and decision-making.

Possible outcomes: Since all the topics mentioned above are related to data management and are relevant to all projects, having a platform for discussion within WG2 allows us to exchange how each of these topics is addressed in all of them. This helps save time, energy, and the efforts of relevant partners dealing with these topics. Thus, it facilitates more efficient collaboration and ensures that best practices are shared and applied across all projects.

3.6 Exposure times of in specific indoor environment

Projects that expressed interest: LEARN, K-HealthAir

Scope & Context: One of the challenges of this kind of projects is how to estimate the exposure levels of people to indoor air pollution according to the time spent in the associated indoor scenarios.

- Accurate estimation of human exposure to air pollutants is mandatory for a realistic evaluation of their risks and for the design and implementation of strategies to control and limit those risks [1].
- Timing of exposure is a key feature for explaining the health effects (of air pollution)[2-3].
- The exposure time depends on time activity profiles of individuals (e.g., time spent in traffic vs. indoors) as well as on the intensity of activity (e.g., sleeping, playing sports) [4].

Regarding this last item, another challenge is how to define and classify the activities. The Multinational Time Use Study (MTUS) brings together more than a million diary days from over 70 randomly sampled national-scale surveys, into a single standardised format. MTUS allows to analyse time spent by people in various sorts of work and leisure activities, over the last 55 years and across 30 countries [5]. Use (simplify) version of the Multinational Time Use Study Activity Code may be a useful tool to later on compare the exposure time of different people, considering not only the time, but also the activity carried out during this period of time.

About how to monitor the personal exposure, there are 2 main strategies:

1. Portable Sensors:
 - a. provide a more reliable “time–geography of exposure”.
 - b. could include physical activity (PA) and location (GPS) tracking sensors.

- c. The implementation of these sensors entails several privacy and ethical concerns related to the collection of personal data.
2. Time Activity Diaries (TAD)
 - a. Preferably Digital TAD (avoid transcription mistakes).
 - b. The participants will be asked to answer a questionnaire about their exposure time to the Air Quality monitoring locations (outdoor, indoor).
 - c. It allows to also collect PA.
 - d. Preferably, to be filled in daily.
 - e. The data can be easily (pseudo)anonymised.
 - f. This option is compatible with portable sensors (to validate the PA sensor measurements).

The best strategy will depend on the nature of the study and the available resources.

Objectives and Actions: Concrete objective and actions still need to be defined.

- To carry out a literature review and evaluate the different strategies used in other studies
- In the case of K-HealthAir we are going to collect indoor and outdoor exposure data through TAD, since the monitoring with sensors was not planned before.

3.7 Harmonization and standardization of data collection techniques across EU labs

Projects that expressed interest: EDIAQI

Scope and Objectives: Consistent standardization of air quality and pollution data collection techniques is essential for accurately forecasting, analyzing, and assessing air quality and thus, improve the impact on public health. In addition to the work on the IPCHEM templates which is described in section 5, the following objectives need to be taken into account when collecting data:

- a clear distinction between different air pollution data, i.e., indoor air pollution such as sensor data, health data, biological data, environmental data, human biomonitoring data, survey data, indoor air sampling data, climate data, and toxological data.
- The types of data collection or data collection methods are based on surveys, interviews, and questionnaires, as provided by IPChem on the one hand, however, we argue that upon considering data collection also sensorics, sampling, and databases need to be taken into account. Therefore, the pollutants and parameters measured via sensorics, such as mobile spirometers, which can be connected to smartphones, stationary pollutant traps, low-cost sensorics (sentinels), air quality monitoring stations, remote reference stations, filtrations systems, such as ventilation and air condition systems, also need to be comprised when standardizing data.
- the venues of data collection require a clearer distinction among these three categories: (1) residential/homes/households, (2) public buildings indoor, and (3) outdoor public venues.
- Most prominently, the data that are measured themselves need to be reported within a standardized chart.

Therefore, EDIAQI project has found a task force to collaborate beyond EDIAQI's four pilots and four campaigns with all seven IDEAL WG2 cluster projects to categorize chemical as well as non-chemical data within an air quality data chart comprising air pollution data definitions within highly structured and less granular categories. The streamlined definitions should result in a more concise reporting and data analysis. Reinforcing the use of this standardized air quality data chart could foster the continuous integration of this scale as new standard within the Horizon Europe Data Management Plan³, thus significantly improve data management within health cluster activities and lead to a clearer and more effective collection, categorization, assessment, and analysis of air quality data.

Timeline: Actions & Outcomes

Tasks & Time	01/2023	07/2023	01/2024	07/2024	01/2025	07/2025
IPChem Data Transfer	Feedback Trial IPChem Forms	Dataset Sharing by Data Owners	Filling of Updated Forms by Data Owners	Dataset Sharing by Data Owners	Filling of Consolidated Forms by Data Owners	Dataset Sharing by Data Owners
Standardization via IDEAL Data Collection Chart	Creating a Data Collection Chart for Chemical and Non-Chemical Data	IDEAL Project Partners fill Chart and provide Feedback	Consolidation of Data Collection Chart based on Partner Answers and Data Management Plan Data Charts	IDEAL Project Partners re-fill Chart	Edited Chart is streamlined within Updated Data Management Plan	
Harmonization of Data Collection Techniques					Updated IPChem Forms and Updated Chart and can be harmonized within Standardized DMP	New Harmonized Standard is presented to EC

Tangible Results	Report to IPChem: Improvement of IPChem Data Quality and Quantity - via feedback on Metadata and Technical Form (EDIAQI) - via Air Quality Data Chart	Report within IDEAL WG2: Streamlined IDEAL Reporting Chart within IDEAL Deliverable	Report to EC: New DMP Standard is reported to EC
-------------------------	---	---	--

As so far only one project has shown interest in this topic, a decision needs to be made whether it will be considered for further discussions within WG2.

3.8 Definition of the hypothesis to be analysed with the data analysis

Projects that expressed interest: K-HealthAir

Scope and Context: A hypothesis in a data-driven study is a guess about a specific parameter or population. Once the hypothesis is defined, the next step is to collect the kind and amount of data sets needed to determine if there is enough evidence to confirm that the hypothesis is true.

Regarding the collection of data, in IDEAL cluster' projects it is important to consider we are going to use personal and health data, and even the IAQ data can be considered sensitive data in some cases. Basically, there are 2 constraints:

- Limitations in available health data: To obtain Ethical Approval, it is essential to adequately justify the type and quantity of data to be collected, as well as the motivation for collecting it, while attempting to restrict the types of data collected. This requirement is outlined in Article 5c of the GDPR and Article 4c of Regulation (EU) 2018/1725, both pertaining to the principles of personal data processing, commonly referred to as data minimisation. Any

addition of new health data must be appropriate for the specific purpose and comply with the aforementioned regulations.

- Limitations in sample size: In studies involving humans, sample size presents a critical ethical concern, compounded by resource constraints. Striking a suitable balance between gathering sufficient data to achieve the study's objectives and minimizing the collection of personal data is always challenging. This balance must be achieved while adhering to the principles of data minimisation.

A way to mitigate both constraints is to carry out a systematic literature review on the study topics. Knowing the scientific evidence around our study topics could orient us to define as precise as possible the hypothesis to be answered and the needed data sets.

Objectives and Actions: Concrete objective and actions still need to be defined. As so far only one project has shown interest in this topic so far, a decision needs to be made regarding whether it will be considered for further discussions within WG2. Defining the hypothesis in an iterative process with all the partners, pursuing the maximum precision in the definition, and checking that the kind and amount of data needed to answer the hypothesis will be available.

4. DMPs across different Projects

4.1 DMP common Timeline

The Gantt chart in Figure 2 illustrates the timeline of DMP activities for each project. Each row represents a specific project, and the time scale is marked horizontally. The chart utilizes different colours to indicate various activities:

- Creation of the Data Management Plan (e.g., yellow): This phase involves the initial development and drafting of the DMP for each project. It includes defining data management goals, outlining data collection and storage methods, and identifying data sharing and documentation practices.
- Refinement of the DMP (e.g., blue): This stage signifies the iterative process of revising and improving the DMP. It includes e.g. incorporating feedback from stakeholders, addressing any identified gaps or challenges, and ensuring compliance with relevant data management policies and regulations.

Establishing data management activities within the project (e.g., green): This category highlights planned activities related to implementing the data management practices outlined in the DMP. It involves developing and setting up data collection systems, organizing data repositories, establishing data quality assurance processes and ensuring data security measures.

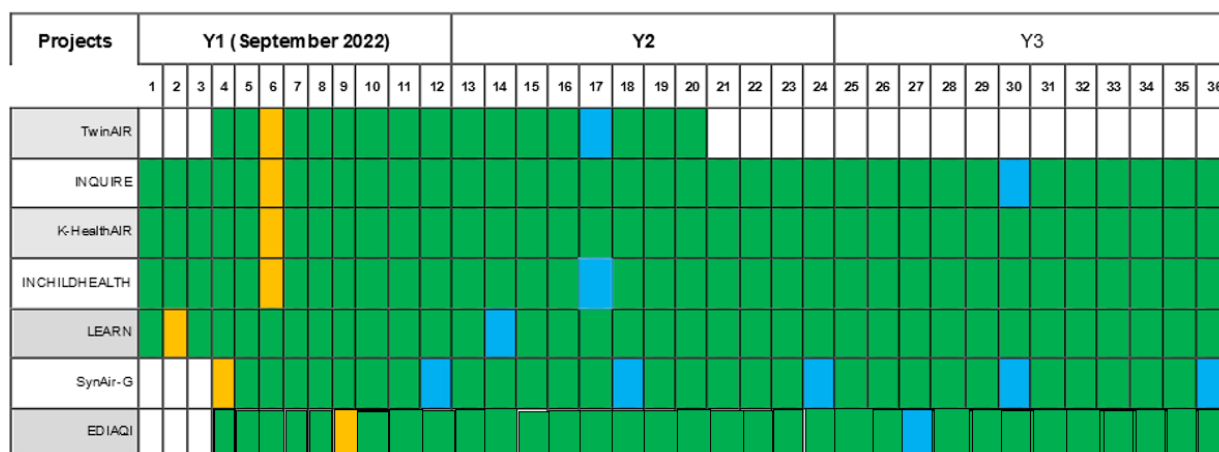


Figure 2 Timeline off all 7 projects considering their DMP development

4.2 DMP collaboration

When it comes to the DMP collaboration within the IDEAL Cluster, all projects have agreed to provide their DMP documentation to be shared and accessed by all projects participating in WG 2 activities. This has been clarified with the project coordinators and will be guaranteed.

5. Collaboration with IPCHEM

5.1 The role of IPCHEM

The Information Platform for Chemical Monitoring Data (IPCHEM) is the European Commission's reference platform for chemical monitoring data collected across various media by the European Commission and Agencies, EU Member States, international and national organisations, EU funded projects and research communities [6]. It supports a coordinated approach for collecting, storing, accessing, and comparing data related to the occurrence of chemicals, and their metabolites, in relation to humans and the environment. IPCHEM data are increasingly used to support the building of the scientific evidence base for the European Union's environment and health policies. IPCHEM facilitates chemical monitoring data sharing between different legal clusters and, therefore, between Member States competent authorities, the Commission services and EU agencies, which is an important factor that influences the effectiveness of the implementation of the EU chemicals legislation.

IPCHEM contains a thematic module "Indoor Air Monitoring". IPCHEM provides harmonised data and metadata templates and code lists related to indoor air monitoring data. Some of the IDEAL projects also collect human biomonitoring (HBM) data. HBM data are integrated by IPCHEM in its module on Human Biomonitoring data.

5.2 Objective and practical organisation

IDEAL WG2 will facilitate the collaboration between the IDEAL projects and IPCHEM to make the relevant datasets collected in the projects, available through IPCHEM in a harmonised way.

In the IDEAL WG2 meeting in February 2023 IPCHEM presented itself and its way of operating in general and specifically for its module "Indoor Air Monitoring". As a result of this meeting the IDEAL-

IPCHEM Data Management Team was set up. It had its first meeting in May 2023. It will convene regularly to organise, follow up and streamline the collaboration between the IDEAL projects and IPCHEM. The IDEAL-IPCHEM Data Management Team consists of participants of all cluster projects and of IPCHEM.

TwinAir, EDIAQI, INQUIRE and LEARN will generate human biomonitoring (HBM) data. The collaboration on these data will be organised in a subgroup of the IDEAL-IPCHEM Data Management Team consisting of the involved projects.

5.3 Identified actions

Identification of relevant datasets that will be shared with IPCHEM

As illustrated in section 2.3 the IDEAL projects will generate a broad variety of data. IPCHEM will integrate quantitative, quality checked data on chemical occurrence, particulate matter (PM) and black carbon (BC).¹ Each IDEAL Project will verify the compatibility of the dataset(s) it generates with the IPCHEM scope and requirements.

Typically, chemical occurrence data in indoor environments are acquired through sampling and lab analysis. The IPCHEM data model has been developed for this type of data based on the substance being measured. The data model also allows to include sum parameters for substances, such as total volatile organic compounds (tVOC), and measurements of PM and BC.

Several projects will also generate chemical occurrence data from sensors. These data will be integrated at a level of aggregation supporting relevant use cases. This level of aggregation will be defined and clarified further. As for all data uncertainties and ranges of uncertainty have to be reported in the lab/analytical part of the metadata.

As mentioned above four projects will generate HBM data. These data will be integrated in the IPCHEM Human Biomonitoring module.

Informed consent forms and agreements

For data from public buildings, such as hospitals, railway stations or schools, the IPCHEM Participation Form [7] allows to indicate who can have access to the data, including public availability. Data can be made available already earlier to EU or national authorities.

The projects TwinAir, EDIAQI, INQUIRE, LEARN, K-HealthAIR and InChildHealth will collect data related to individual homes. Data on individual homes (including measurements in the home, or from samples taken in the home) are personal data. The projects TwinAir, EDIAQI, INQUIRE and LEARN, SynAir-G will also collect HBM data which are personal sensitive data. Integration of personal data in IPCHEM has to comply with the GDPR/ EUDPR [8]. This requires an informed consent (IC) form to be signed by the participants to the studies (data subjects). To allow integration of the (pseudonymised) individual data in IPCHEM the IC form has to explicitly mention sharing of the data for (re)use beyond the project goals.

The latest available version of the IC form template drafted in the Partnership for the Assessment of Risks of Chemicals (PARC) [9] has been shared with the IDEAL projects. This template mentions IPCHEM explicitly. However, in some of the projects IC forms have already been signed before this version of the PARC template was shared. In the INQUIRE project a more general phrasing was used based on the status of the PARC IC form template at the moment of drafting the INQUIRE informed consent form. IPCHEM will verify the IC forms already used in the projects.

¹ Biological agents (targeted measurements) are included in the indoor air data template. However, considering that IPCHEM will be integrated into the Common Data Platform on Chemicals in the future, it is not guaranteed that biological data will be transferred. Therefore, they are not included here, although projects that would like to include them, have this possibility.

For integrating (pseudonymised) personal data in IPCHEM joint controller agreements need to be signed by IPCHEM and the data controllers in the projects. In PARC a proposal for a joint controller agreement is actually under discussion. Comments from PARC are expected by early September. Afterwards the joint controller agreement will be shared, discussed and, if necessary, amended within the IDEAL cluster.

If integration of single measurement data is not possible, the data will be aggregated to a level where identification of data subjects is no longer possible.

Harmonisation of data and metadata

As far as possible common templates for data and metadata will be used by the IDEAL projects for the data that will be integrated in IPCHEM. The starting point will be the metadata and data templates developed by IPCHEM for the IPCHEM thematic module 'Products and Indoor Air Data' [10] IPCHEM provided an excel template for supplying metadata. The IDEAL projects will fill in this template for their planned studies in order to identify any needs for adaptation. This will be discussed and decided in the meetings of the IDEAL-IPCHEM Data Management Team, upon which IPCHEM will provide an adapted template. The goal is to have the harmonised template ready by the end of 2023 after which projects will upload their draft metadata. These metadata will be updated later to reflect actual data integration.

In a similar way the data template will also be evaluated for harmonisation and use by the IDEAL projects.

For the HBM data the templates developed in the HBM4EU project [11] and further elaborated in PARC will be used.



6. References

- [1] Air Pollution, the Automobile, and Public Health. Watson AY, Bates RR, Kennedy D, editors. Washington (DC): National Academies Press (US); 1988.
- [2] <https://www.epa.gov/report-environment/outdoor-air-quality>
- [3] Y. Sellier et al., Health effects of ambient air pollution: Do different methods for estimating exposure lead to different results?, Environment International, Volume 66, 2014, Pages 165-173
- [4] Sarigiannis, D., Chapizanis, D., Arvanitis, A. (2018) Report on the methodology for estimating individual exposure, Deliverable D4.1, ICARUS H2020 Project
- [5] <https://www.timeuse.org/mtus>
- [6] Kephelopoulou, S., Bopp, S. K., Costa, S. D., Cusinato, A., Lipsa, D., & Geiss, O. (2020). Indoor air monitoring: Sharing and accessing data via the Information Platform for chemical monitoring (IPCHEM). International Journal of Hygiene and Environmental Health, 227, 113515. <https://doi.org/10.1016/j.ijheh.2020.113515>
- [7] Annex 1 of the IPCHEM Data Policy (https://ipchem.jrc.ec.europa.eu/documents/IPCHEM_DataPolicy_V3.pdf)
- [8] GDPR: General Data Protection Regulation (Regulation (EU) 2016/679). EUDPR: As IPCHEM is hosted by the European Commission, the EUDPR (Regulation (EU) 2018/1725 on the protection of natural persons with regard to the processing of personal data by the EU institutions, bodies, offices and agencies and on the free movement of such data) is the corresponding data protection legislation that applies to EU institutions.
- [9] <https://www.eu-parc.eu/>
- [10] Kephelopoulou, S., Bopp, S. K., Costa, S. D., Cusinato, A., Lipsa, D., & Geiss, O. (2020). Indoor air monitoring: Sharing and accessing data via the Information Platform for chemical monitoring (IPCHEM). International Journal of Hygiene and Environmental Health, 227, 113515. Supplementary Material SM1. (<https://ars.els-cdn.com/content/image/1-s2.0-S1438463919310752-mmc1.docx>) and SM2 (<https://ars.els-cdn.com/content/image/1-s2.0-S1438463919310752-mmc2.docx>)
- [11] <https://www.hbm4eu.eu/>
-