



Deliverable D4.5

GitHub Repository of EDIAQI models Code - Version 1

Work Package 4

MONITOR – Pilots, data and campaigns

Version: Final

Deliverable Overview

This deliverable provides an overview of the open-source EDIAQI GitHub repository that contains code used within the project. The EDIAQI GitHub repository will be continuously enhanced by partners throughout the project lifecycle.

Additional Information

Type: Other

Dissemination Level: PU – Public

Official Submission Date: 30th of November 2024

Actual Submission Date: 26th of November 2024

Disclaimer

This document contains material, which is the copyright of certain EDIAQI partners, and may not be reproduced or copied without permission. All EDIAQI consortium partners have agreed to the full publication of this document if not declared “Confidential”. The commercial use of any information contained in this document may require a license from the proprietor of that information. The reproduction of this document or of parts of it requires an agreement with the proprietor of that information.

Document Revision History

Version	Date	Description	Partners
V0.1	11 th October 2024	First draft of the deliverable	KNOW
V0.2	13 th November 2024	Second draft of the deliverable	KNOW
V0.3	18 th November 2024	Final draft of the deliverable	KNOW
V0.4	25 th November 2024	Internal review of the deliverable	SCH + FTMC
V1.0	25 th November 2024	Update after the internal review	KNOW
Final	26 th November 2024	Final quality check	LC

Authors and Reviewers

Authors

- Kristina Časni (KNOW)
- Heimo Gursch (KNOW)

Reviewers

- Ivana Banić (SCH)
- Steigvilė Byčenkienė (FTMC)

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

List of Figures.....	5
Executive Summary	6
1. Introduction	7
2. Consortium Overview and Contributions.....	8
3. GitHub Structure	10
4. Submission and Review Process	11
5. Conclusion.....	12

List of Figures

Figure 1. Repositories of the EDIAQI GitHub.	8
Figure 2. Screenshot of the EDIAQI profile on GitHub.....	10

Executive Summary

The aim of this deliverable is to provide a brief overview of the EDIAQI GitHub page under the GitHub name EDIAQI and its structure. The page itself can be accessed by following the link below. In the deliverable we provide information on the structure and what visitors can expect to see in each of the repositories and sub-repositories. The EDIAQI GitHub will form the central access point for open-source model code that is and will be the outcome from the project. The repository will be continually enhanced and expanded as the project advances.

[Access the EDIAQI project's GitHub](#)

1. Introduction

The EDIAQI project intends to share as many details as possible of the executed work within the project, following open science practices to make research results accessible and reusable for both scientists and the general public. One of the practices implemented is publishing work as pre-prints and in open access journals. Data will be accessible through IPCHEM, and in cases where data publishing is not possible due to GDPR, ethical, commercial or any other reason, this will be clarified in the metadata description. The project's research data will be stored on the KNOW Data Platform (as described in Deliverable D4.1), making (meta) data publicly available and accessible. A third aspect of the project's open science practices is making the code used within the project accessible and well documented which is done through GitHub. The goal of this open science approach is to stimulate further research by other scientists, who were not included in the project, after the project ends. It is also for the project partners to continue work and further development, contributing all together to better understanding of the indoor air pollution and its effects.

2. Consortium Overview and Contributions

A total of eighteen partners are involved in the EDIAQI project, with work distributed across seven different work packages that collaborate closely with one another. The resulting code from the project will be made publicly available in five different repositories in the GitHub profile of the project named EDIAQI, each representing a distinct group of the EDIAQI project's work as seen on the Figure 1.

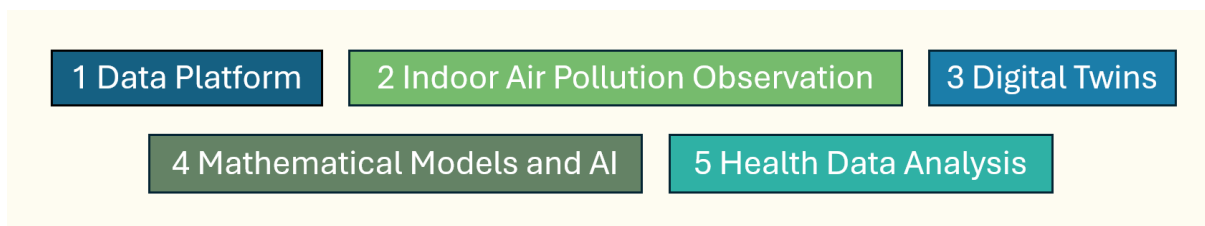


Figure 1. Repositories of the EDIAQI GitHub.

Three of the five mentioned repositories are part of Work Package 4 (WP4), including the Data Platform, Digital Twins, and Mathematical Models and AI. The content of each repository depicted in Figure 1 follows the organisation of the EDIAQI project into five repositories:

- **Data Platform - Task 4.1** - which focuses on the development of the Data Platform, is led by KNOW partner, collaborating with LC, ASC, THIN, DEDA, TalTech, and LAS. The respective GitHub repository contains the code necessary for the development of the platform, which is used within the project. The Data Platform was developed for data-driven collaboration within the consortium to ensure FAIR data management, secure data sharing, and reproducible analytics.
- **Indoor Air Pollution Observation** - The Indoor Air Pollution Observation is part of Work Package 3 (WP3) and is led by TROPOS with support from IMROH, ANT, TUG, UMOL, TalTech, LAS, and FTMC. The Air Quality Sensor Evaluation Toolkit is part of Task 3.1, which aims to identify the most suitable approaches, tools and methods to characterize indoor air pollution and to estimate concentration levels thereof with the deployment of user-friendly technologies tested at TROPOS laboratories in parallel to real-life experiments. This main repository also contains a sub-repository named microbiome which contains basic data analysis workflow for 16S rRNA gene

fragment and ITS region amplicon sequencing by TUG. More details and the code can be found in the respective GitHub sub-repository.

- **Digital Twins** - The Digital Twins repository is currently empty at the time of this deliverable's submission, but it will contain the code produced from Task 4.4, titled Hybrid approach for digital twins, where ASC is building physics-based models using data from the pilots, collaborating with partners TROPOS and KNOW. The goal of this task is to create simulated environments that are an adequate representation of the physical world to better understand indoor air quality.
- **Mathematical Models and AI** - The final group of WP4, Mathematical models and AI, as part of Task 4.5, is led by KNOW and focuses on applying data analysis and machine learning to the project's data. This work is carried out in a collaboration with partners ASC, WINGS, TROPOS, TalTech, LAS, and NIB. As part of this task, the Simulation Toolbox has been developed by RegionH and KNOW to provide a public tool for risk assessment given location, surroundings, building type, and information on inhabitants' information. Another part of the task which is currently available is applied machine learning for low-cost sensors by ASC. More details, along with the code, can be found in the respective GitHub sub-repository.
- **Health Data Analysis** - The Health Data Analysis repository contains code for work being performed within Work Package 5 (WP5), which aims to assess both the short- and long-term effects of indoor air pollution on human health. This is achieved through Task 5.1, by analysing data of retrospective cohorts (led by RegionH and supported by KNOW, TUG, and SCH), Task 5.2. by analysing prospective cohorts (led by SCH with contributions from KNOW, IMROH, ASC, WINGS, RegionH, ANT, TUG, and TROPOS), and Task 5.3, where toxicological studies are done by IMROH as the lead with support from RegionH, UMOL, and NIB. At the time of submission, the analysis of the Atopica and COPSAC2000 cohorts are available.

More information about the project, its consortium, and work packages can be found on the [EDIAQI Project Website](#).

3. GitHub Structure

The EDIAQI project's GitHub page currently contains five main repositories as shown in Figure 2. Each of these repositories collects open-source code related to the work done in each task represented by the repository name.

For additional clarity, README.md files are present in each sub-repository. The text from each of these files is displayed to the users by default when they navigate to an individual sub-repository to detail the work done in that task so far, how the current code is organised, and which tools were used for code production (e.g. Python, R, etc.).

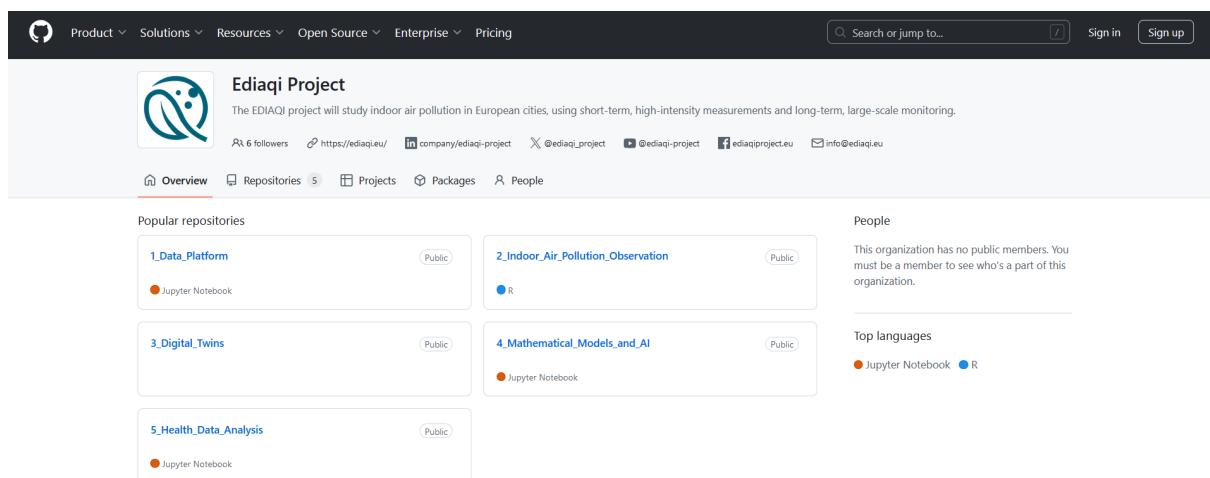


Figure 2. Screenshot of the EDIAQI profile on GitHub

4. Submission and Review Process

Some of the eighteen partners who are part of the EDIAQI project contribute with different analyses, tools and developed models, resulting in code that can be publicly shared via GitHub. The EDIAQI GitHub collects all the code from the EDIAQI project that can be publicly shared.

Only approved members of the EDIAQI GitHub can submit their work, and the process is as follows:

- Partners develop code, models, tools, etc.
- Partners have internal development guidelines for quality assurance.
- Code is shared within the EDIAQI GitHub, but it is not available to the public during code reviewing and testing:
 - The review is done from partners in the same work package, who are familiar with the requirements, but did not work on the code themselves.
 - The partners must agree that the code fulfils the EDIAQI guidelines before it is published. This process is inspired by the review process of the deliverables.
- Code is made available to the public via GitHub once it is finally tested and reviewed.

API keys and other confidential information, such as passwords, should not be shared via GitHub. To address this, the solution in EIDAQI is to create a separate file that holds all confidential information of the work, typically as a YAML or properties file. The sensitive information is then removed from the file, so that only the names of the sensitive information remain inside, and the file can then be submitted to the GitHub. In the README.md file the content of the sensitive file should be explained, so that users can easily insert there their own API keys, passwords, etc.

During the review process, the partners check if the handling of sensitive information was done correctly.

5. Conclusion

The GitHub repository contains the open-source code generated during the project's duration, structured to be intuitive for external users who were not part of the EDIAQI project. Currently, open-source code for the work executed so far in the project is available. This will continue to evolve as the project progresses and will be finalised in M32 with the second version of Deliverable D4.5, titled GitHub Repository of EDIAQI models Code - Version 2.

The open-source code is divided among five tasks, organised into five main repositories within the EDIAQI project's GitHub profile named EDIAQI. Processes for reviewing the code on the GitHub repository have been established.

There may be changes in the GitHub repository structure as the project evolves, if the need for better organisation and clarity occurs. All modifications will be documented and explained in the second version of Deliverable 4.5, which is due in M32.

All repositories will be updated as needed as the work within the project progresses. Additional code will include the code of the digital twins, toxicological studies, analyses of prospective cohorts, project data in general and machine learning applications.



Deliverable D4.5

GitHub Repository of EDIAQI models Code - Version 1

Work Package 4

MONITOR – Pilots, data and campaigns

Version: Final



This project has received funding from the European Union's
Horizon Europe Framework Programme under grant agreement N° 101057497.