

REENGAGE

ENGAGING CITIZENS - MOBILIZING TECHNOLOGY - DELIVERING GREEN DEAL

REENGAGE experiment and monitoring system and manual 2

Work Package 4, Deliverable D4.6

D4.6 GREENGAGE experiment and monitoring system and manual 2

Work package 4, Deliverable D4.6

Please refer to this report as follows:

Sánchez, R., Emaldi, M., Vergara, F., Lopez de Ipiña Gonzalez de Artaza, D., et al. (2025). GREENGAGE experiment and monitoring system and manual 2. Deliverable D4.6 of the Horizon Europe project GREENGAGE.

Project information	
Project name:	GREENGAGE – Innovative governance, environmental observations and digital solutions in support of the Green Deal
Grant Agreement No.	101086530
Start date:	01/01/2023
Duration:	36 months
Coordinator:	AIT Austrian Institute of Technology Giefinggasse 4, 1210 Vienna, Austria
Contact:	Jan Peters-Anders, Research Engineer
	Funded by the European Union GA n° 101086530.
Deliverable details	
Description:	This report covers T4.3 of project GREENGAGE and provides a detailed list of implemented features for the system, their testing and how to use them. It also provides details on how to access various statistics for impact assessment.
Version:	Final
Dissemination level:	PU (Public)
Due date:	30/06/2025
Submission:	04/07/2025
Lead:	University of Deusto
Author(s):	Sánchez, R., Emaldi, M., Vergara, F., Lopez de Ipiña Gonzalez de Artaza, D., (DEUSTO), Spain Wolosiuk D. (VRVIS), Austria Khrystodulov A. (VISAT), Czech Republic Martens S. (BUAS), The Netherlands Lizama A. (MindEarth), Switzerland Pirker M. (SushiDev), Austria

	<p>Covato E. (UWE), UK</p> <p>Alabyan D. (Digital Sunray), Austria</p> <p>Peters-Anders J., Kozłowska A., Gebetsroither-Geringer E. (AIT), Austria</p>
--	--

AWAITING VALIDATION BY THE EUROPEAN COMMISSION

Legal Disclaimer

All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

Users wishing to reuse material from this work that is attributed to a third party, such as tables, figures or images, are responsible for determining whether permission is needed for that reuse and for obtaining permission from the copyright holder. The risk of claims resulting from infringement of any third party-owned component in the work rests solely with the user.

© 2025 by GREENGAGE Consortium

Table of Contents

Content

1	GREENGAGE summary	2
2	Introduction	3
3	GREENGAGE Citizen Observatory Community Journey	5
3.1	Area of Concern 1: Community and Co-production Process Management	9
3.1.1	Collaborative Environment.....	9
3.1.2	Discourse	13
3.1.3	WordPress	17
3.2	Area of Concern 2: Data Crowdsourcing and Curation	20
3.2.1	MindEarth for GREENGAGE app	20
3.2.2	GREENGAGE app	24
3.2.3	MODE	26
3.2.4	IoT Sensors.....	27
3.3	Area of Concern 3: Analysis and Visualization for Insights Generation	28
3.3.1	Apache NiFi	29
3.3.2	Apache Druid.....	31
3.3.3	Apache Superset	34
3.3.4	Data Quality Dashboard	36
3.3.5	UrbanTEP / VISAT	37
3.3.6	DataHub	40
3.3.7	DigitalTwin.....	42
3.3.8	Multi-modal data processing tools.....	44
3.4	Horizontal services.....	45
3.4.1	Keycloak.....	45
3.4.2	Zenodo	48
4	Usage manual of GREEN Engine and GREENGAGE's Academy	51
4.1	Overview of the Co-production Framework	51
4.2	Setting up a Thematic Co-exploration.....	52
4.2.1	Preparing the Campaign in the Collaborative Environment.....	52
4.2.2	Stakeholder Onboarding with Discourse and Collaborative Environment.....	52
4.3	Designing the Crowdsourcing Campaign	53
4.3.1	Configuring POIs and Tasks.....	53
4.3.2	Deploying the GREENGAGE App	54
4.4	Data Collection and Transformation	54
4.4.1	Executing the Campaign.....	54
4.4.2	ETL and Data Integration	54
4.5	Data Analysis and Visualization	55

4.5.1	Creating Dashboards with Apache Superset.....	55
4.5.2	Policy Dashboard and Reporting Templates.....	57
4.6	Feedback, Impact and Dissemination.....	58
4.6.1	Policy and Social Engagement.....	58
4.6.2	Impact Assessment and Zenodo Publication.....	59
4.7	Conclusions and Recommendations	60
5	Usage statistics & validation	62
5.1	Collaborative Environment.....	62
5.2	Discourse.....	63
5.3	WordPress.....	65
5.4	MindEarth for GREENGAGE app.....	67
5.5	GREENGAGE app	67
5.6	MODE.....	68
5.7	IoT Hyperlocal Sensors	68
5.8	Apache NiFi.....	69
5.9	Apache Druid.....	69
5.10	Apache Superset.....	71
5.11	Data Quality and Structure Dashboard	72
5.12	UrbanTEP / VISAT	72
5.13	DataHub.....	73
5.14	DigitalTwin.....	74
5.15	KeyCloak	74
6	Conclusions.....	77

List of Figures

Figure 1: Structure of project GREENGAGE.....	2
Figure 2: Overview of the Citizen Observatory Community Journey	8
Figure 3: Example of instantiation of CO Enablers (Citizen Observatory's Thematic Co-Exploration specification template).....	10
Figure 4: View in Collaborative Environment of the reference co-creation process model for Thematic Co-Explorations.....	12
Figure 5: Reference process model created to govern Thematic Co-Explorations' co-creation.....	13
Figure 6: Landing wizard for new users.....	15
Figure 7: A Discourse topic in which the automatic translation button can be seen.	16
Figure 8: Result of the automatic translation of a Discourse topic.....	16
Figure 9: Discourse API Documentation.....	17
Figure 10: Mailchimp newsletter subscribing form.	20
Figure 11: MindEarth for GREENGAGE app running on smartphones.	21
Figure 12: GraphQL view of the GREENGAGE app admin console.....	24
Figure 13: GraphQL view of the GREENGAGE app.	25
Figure 14: Smart Spot device.....	27

Figure 15: Libelium One Sound Level device.....	27
Figure 16: Atmotube PRO wearable sensor	28
Figure 17: Example of a NiFi dataflow.....	30
Figure 18: Apache Druid data source view.....	32
Figure 19: Apache Druid analytics view.	33
Figure 20: Example of Apache Superset Dashboard.	34
Figure 21: SQL lab from Apache Superset.....	35
Figure 22: Data Quality Dashboard.	37
Figure 23: Examples of GISAT’s web-based analytical applications with VISAT framework application.	38
Figure 24: Overall technical architecture of the web application.....	40
Figure 25: Example of upgraded UrbanTEP’s VISAT GUI interfaces.....	40
Figure 26: Explore view of DataHub.....	41
Figure 27: An example of an analysis in the Dashboard	43
Figure 28: User view of the socio-demographic form.....	46
Figure 29: Keycloak’s administrator view of socio-demographic data.	47
Figure 30: Citizen Science Loop in GREENGAGE, starting from “Problem Identification” and finishing in “Action&Impact”.	51
Figure 31: GREENGAGE Collaborative Environment	53
Figure 32: GREENGAGE app’s backend dashboard	53
Figure 33: Three GREENGAGE app’s screenshots used whilst observing POIs at Deusto University’s campus.	54
Figure 34: Apollo Server’s GraphQL API.	55
Figure 35: Apache Druid’s data store for GREENGAGE.	55
Figure 36: Per POI visualization of air quality perception and pollution detection via Atmotube.	56
Figure 37: Ad-hoc visualization showing pictures taken by observers at different spots in Deusto’s campus.	57
Figure 38: Discussion forum for Deusto’s campus’ thematic co-exploration.	58
Figure 39: Policy brief summarising recommendations for University of Deusto’s campus’ administrators.	59
Figure 40: Impact analysis performed following the ACTION project’s impact evaluation approach.	60
Figure 41: Datasets created in the Thematic Co-Exploration, uploaded into GREENGAGE’s Zenodo community.	60
Figure 42: Technical documentation page where the step-wised co-production of a Thematic Co-Exploration following GREENGAGE’s approach and using its tools and knowledge assets is detailed.....	61
Figure 43: Chart depicting the processes created through Collaborative Environment from its beginning to June 2025.	62
Figure 44: Chart depicting the users who have registered/entered for first time in the Collaborative Environment.....	63
Figure 4546: Chart depicting the teams created in the Collaborative Environment.	63
Figure 4748: An example of analytics built into Discourse deployment for GREENGAGE.....	64
Figure 49: Different statistics gathered at Discourse. Topics created (upper left), posts created (upper right), pageviews (bottom left) and time to first response (bottom right).....	65
Figure 50: overview of the statistics from the current year.....	66
Figure 51: Distribution of visits from social networks.....	66
Figure 52: Mission control records for testing missions.	67
Figure 53: Example of interactions between MQTT Clients through an MQTT Broker.	69
Figure 54: Dashboard View of Apache Druid.	70
Figure 55: Task View of Apache Druid.....	70
Figure 56: Superset Chart view.	71
Figure 57: Superset Dataset view.....	72

Figure 58: DataHub Analytics panel view.....	73
Figure 59: DataHub entities per platform count.....	73
Figure 60: Policy, data and indicator flow chart in DigitalTwin.....	74
Figure 61: Authentication errors registered by KeyCloak deployment of GREENGAGE.....	75
Figure 62: Clients submitting authentication requests to GREENGAGE's KeyCloak-based authentication service.....	75
Figure 63: Distribution of errors in GREENGAGE's KeyCloak-based authentication service.....	76

List of Tables

Table 1: GREENGAGE components summary.....	49
Table 2: Summary of published campaigns for testing	67

List of Acronyms

APK	Android Package Kit
CE	Collaborative Environment
CO	Citizen Observatory
ETL	Data integration process Extract, Transform, Load
HDFS	Hadoop Distributed File System
HTTP	Hypertext Transfer Protocol
JDBC	Java Database Connectivity
JOLT	JSON Language for Transform
JSON	Java Script Object Notation
KPI	Key Performance Indicators
MQTT	MQ Telemetry Transport

Glossary

Citizen Observatory	According to EU project WeObserve, "Citizen Observatories (COs) are community-based environmental monitoring and information systems, that invite individuals to share observations, typically via mobile phone or the web. Throughout these activities citizens become able to participate in environmental management/local governance." [Reference: D2.1 - GREENGAGE Methodological Framework]
Citizen Science	In Citizen Science, a broad network of people collaborates. Participants provide experimental data and facilities for researchers, raise new questions and co-create a new scientific culture. While they add value, volunteers acquire new learning and skills and gain a deeper understanding of the scientific work in appealing ways. As a result of this open, networked and transdisciplinary scenario, science-society-policy interactions are improved, leading in turn to a more democratic research, based on evidence and informed decision-making" (Socientize consortium, 2014, p. 10).
GREEN Engine	GREENGAGE's dedicated technological platform leveraging GREEN Infrastructure to enable GREENGAGE CO activities.

Executive summary (publishable)

The document "*D4.6 GREENGAGE experiment and monitoring system and manual*" from the Horizon Europe project GREENGAGE focuses on innovative governance, environmental observations, and digital solutions supporting the Green Deal. This document's audience is any of the stakeholders that will be involved in a Thematic Co-Exploration and any person who aims to understand the Citizen Observatory Community Journey. It details citizen engagement in environmental monitoring through Citizen Observatories in European cities. The document outlines three areas of concern: Citizen Observatory Community Journey, including Community and Co-production Process Management, Data Crowdsourcing and Curation, and Analysis and Visualization for Insights Generation. It emphasizes the use of various digital tools and applications for data collection, analysis, and visualization, aiming to enhance city decision-making processes and governance with citizen participation and diverse data integration.

Additionally, the document includes a comprehensive review of the usage statistics and impact assessment for each digital tool and application used in the GREENGAGE project. This analysis focuses on evaluating the effectiveness and reach of these tools in engaging citizens, enhancing data collection, and contributing to better environmental governance and decision-making processes in European cities. The assessment aims to measure the success of the project in terms of citizen engagement and the quality of data obtained through these innovative digital means. This document is a second iteration of deliverable D4.2. Changes from the previous version are clearly indicated.

Related documents

Deliverable D4.1 – GREEN Engine and manual 1

Deliverable D4.2 – GREEN Engine and manual 2

Deliverable D4.9 – GREEN Engine and manual 3

Deliverable D2.6 – GREENGAGE technological requirements 1

Deliverable D2.7 – GREENGAGE technological requirements 2

Deliverable D4.5 – GREENGAGE experiment and monitoring system and manual 1

1 GREENGAGE summary

The pan-European Innovation Action, funded under the Horizon Europe Framework Programme, aims to promote innovative governance processes, and help public authorities in shaping their climate mitigation and adaptation policies. To achieve this aim, the GREENGAGE project will leverage citizens' participation and equip them with innovative digital solutions that will transform citizen's engagement and cities' effectiveness in delivering the European Green Deal objectives for carbon neutral cities.

Focusing on mobility, air quality and healthy living, citizens will be inspired to observe and co-create their cities by sensing their urban environments. The aim is to complement, validate, and enrich information in authoritative data held by the public administrations and public agencies. This will be facilitated by engaging with citizens to co-create green initiatives and to develop Citizen Observatories (CO). In GREENGAGE, Citizen Observatories will be a place where Pilot cities will co-examine environmental issues integrating novel bottom-up process with top-down perspectives. This will provide the basis to co-create and co-design innovative solutions to monitor environmental problems at ground level with the help of citizens.

With two interrelated project dimensions, the project aims to enhance intelligence applied to city decision-making processes and governance by engaging with citizen observations integrated with Copernicus, Global Earth Observation System of Systems GEOSS, in-situ, and socio-economic intelligence, and by delivering innovative governance models based on novel toolboxes of decision-making methodologies and technologies.

The envisioned Citizen Observatory campaigns are being deployed and fully demonstrated in 5 Pilot engagements in selected European cities and regions including: Bristol (the United Kingdom), Copenhagen (Denmark), Turano and Gerace (Italy) and the region of North Brabant (the Netherlands). These innovation Pilots aim to highlight the need for smart city governance by promoting citizen engagement, co-creation, as well as gathering new data which will complement existing datasets and evidence-based decision and policymaking.

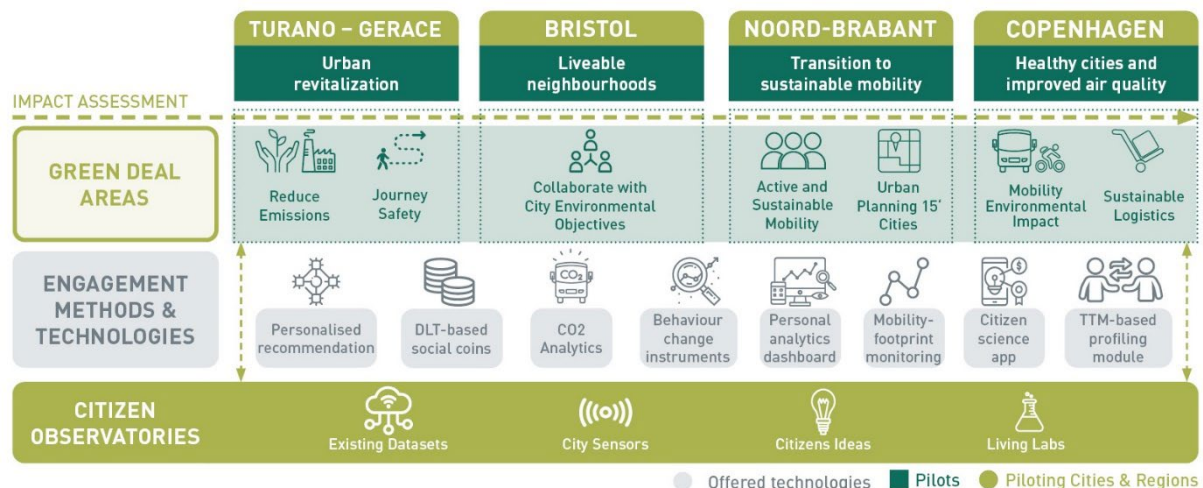


Figure 1: Structure of project GREENGAGE.

2 Introduction

The GREENGAGE project, an ambitious initiative under the Horizon Europe Framework Programme, is at the forefront of promoting innovative governance through the active engagement of citizens in Citizen Observatories (CO). With a mission that intertwines diverse disciplines, GREENGAGE encapsulates a multidisciplinary approach, combining expertise from Engineering and Technology, Computer Science, Social Science and Urban Science. This rich tapestry of knowledge is essential in shaping a project that not only addresses environmental challenges but also resonates with the societal and technological facets of urban living.

Central to this endeavour is Work Package 4 (WP4), titled *Citizen Observatory enabling infrastructure and interoperable toolbox*. WP4 is a critical component of the project, emphasising the Engineering, Technology, and Computer Science aspects of the GREENGAGE solutions. It is here that the project's technological work is found, contributing an assortment of assets crucial for enabling data value chains through CO activities. These activities are designed to inform and validate urban policy design, bridging the gap between grassroots data collection and high-level policymaking.

Task 4.3, nested within WP4, is pivotal in realising the project's ambitions. It focuses on creating a manual that describes the Citizen Observatory Community Journey within the GREENGAGE project. This journey implies the use of the GREENGAGE Experiment and Monitoring System, which is a collaborative system that leverages the co-creation environment established in the INTERLINK project. This system, characterized by an intuitive frontend, is a gateway for defining CO initiatives, guiding experiment execution, fostering team collaborations, and enabling the collection and uploading of experimental data. It stands as a testament to the project's commitment to creating a platform where technology and community engagement converge seamlessly.

In this deliverable (D4.6), titled GREENGAGE Experiment and Monitoring System and Manual 2, we delve into the intricacies of this system. The deliverable is structured to provide a comprehensive guide through the journey of a stakeholder through GREENGAGE-provided resources. It showcases how the project's partners have synergized their varied technological and data-related resources to craft an interoperable suite of tools as a result of iteration 2. This suite forms the backbone of the GREENGAGE infrastructure, designed to support the entirety of the urban analytics data-to-policy workflow lifecycle. Notably, this second version of this deliverable highlights the final version of the GREEN Engine made public for experimentation across pilots. Specifically, a new section has been defined, namely, "4.Usage manual of GREEN Engine and GREENGAGE's Academy", where an example is given on how to use the toolset and approach provided by GREENGAGE, and usage metrics for all GREEN Engine tools have been collected and reported in "5.Usage statistics & validation". Besides, all the descriptions of components encountered in section "3.GREENGAGE Citizen Observatory Community Journey" have been revisited and extended when needed, to accommodate the new developments during iteration 2.

Adhering to stringent ethical and data protection guidelines as outlined in the D1.10 Ethical Management and Activities and D1.12 Data Management Plan of Work Package 1 and D6.3 Privacy Impact Assessment of Work Package 6, the system ensures that the security of data is upheld. This adherence is not just a compliance measure but a core value of the GREENGAGE project, reflecting its commitment to responsible and respectful handling of information.

Central to this system is the facilitation of a dynamic interaction between citizen observers and a consortium of experts in fields like climate and pollution. The platform is enriched with educational resources, including videos and tutorials, to empower citizens with the knowledge and skills necessary for meaningful participation in environmental monitoring and analysis. This interactive and educational approach demystifies scientific processes, making them accessible and engaging for the broader community.

The deliverable delineates the versatility of the system in accommodating diverse Pilot-specific needs. Each Pilot site is enabled to design and implement unique Citizen Science experiments, establishing tailored data workflows, analytical processes, and relevant Key Performance Indicators (KPIs). These KPIs are not mere metrics; they are tools for gauging the impact and effectiveness of each initiative, feeding into the comprehensive evaluation process of Work Package 6.

An innovative aspect of the system is its capability to transform raw data and observations into coherent and compelling narratives. The platform's visualization tools and storytelling features not only aid in data interpretation but also in communicating the insights gleaned in a manner that resonates with the public.

As this introduction unfolds, it sets the stage for a detailed exposition of the three areas of concern that define the Citizen Observatory Community Journey within the GREENGAGE project: *Community & Co-production Process Management*, *Data Crowdsourcing & Curation* and *Data Analysis & Insights Generation*¹. These areas represent the holistic and iterative process of community building, co-production process arrangement, data generation, and analytical exploration. They are not mere steps in a procedure but are the pillars supporting the overarching goal of the project – to engage citizens actively in the pursuit of sustainable urban living and environmental stewardship.

In essence, this deliverable does not merely serve as a manual; it is a narrative of innovation and collaboration. It is a guide that beckons policymakers, environmentalists, technologists, and citizens to embark on a collective journey towards more sustainable and participatory collaborations. As readers delve into the pages of this deliverable, they are invited to envision their role in this transformative project.

¹ Extensively documented at https://greengage-project.github.io/Documentation/thematic_coexploration_example/

3 GREENGAGE Citizen Observatory Community Journey

The GREENGAGE project, a pioneering initiative in citizen-driven environmental monitoring and policy shaping, embarks on a transformative journey known as the *GREENGAGE Citizen Observatory Community Journey*. This journey is a structured pathway designed to empower all the stakeholders that are involved in GREENGAGE to actively engage in Thematic Co-Explorations and decision-making processes mediated by the tools provided by the project. It unfolds in three distinct yet interconnected areas of concern, each integral to the overall success and impact of the GREENGAGE campaigns. Notice that the following technical documentation page details further every component described under the three areas of concern listed below:

https://greengage-project.github.io/Documentation/thematic_coexploration_example/

Area of Concern 1: Community and Co-production Process Management

The tools in the “*Community and Co-production Process Management*” area of concern are not isolated entities but integral parts of a cohesive system. Their interaction is pivotal for the success of subsequent areas of concern.

Collaborative Environment (CE): This digital platform is the nucleus of the project’s community engagement. It enables users to initiate and develop environmental Citizen Science (CS) observation campaigns by providing tools for planning, resource allocation, and campaign maintenance. The environment is envisioned to be user-friendly and robust, facilitating seamless interaction and information exchange. It hosts a resource catalogue, offering guidance and examples to inspire and inform users. The goal is to create a self-sustaining community where knowledge and experiences are shared openly.

Discourse: As a complementary tool to the CE, Discourse serves as a centralized forum for discussion, problem-solving, and idea exchange. It is expected to act as a community pulse, reflecting the concerns, suggestions, and feedback of participants. This platform will be crucial for maintaining transparency, addressing queries, and fostering a sense of community among participants across different Pilot projects.

WordPress: Within the GREENGAGE project, WordPress serves as the main platform for the official website (www.greenage-project.eu), acting as the central public interface for all project-related activities, updates, and resources. The WordPress site facilitates easy access to information about the project’s objectives, pilot initiatives, and opportunities for stakeholder involvement. Integrated with additional tools such as Mailchimp and Matomo Analytics the WordPress site supports real-time analytics, community-building features, and dynamic content delivery. While Discourse fosters dialogue and the Collaborative Environment ensures structured co-production workflows, the WordPress site amplifies outreach and enhances the project’s public impact through effective dissemination and digital presence.

Area of Concern 2: Data Crowdsourcing and Curation

Data Crowdsourcing and Curation encompasses concrete data collection, curation and governance activities. This area of concern is characterised by active participation, leveraging technology to gather vital environmental data.

MindEarth for GREENGAGE app²: This tool enables users to capture detailed, real-time environmental data. Through its advanced mapping and imaging capabilities, MindEarth provides high-quality, geo-referenced information crucial for understanding local environmental conditions. Users can contribute by collecting data on their surroundings, thus playing a direct role in building a comprehensive environmental database.

GREENGAGE app: Designed to be intuitive and engaging, the GREENGAGE app allows citizens to participate in data collection through their mobile devices. This app includes various features for data collection, tailored to suit the needs of different Pilot campaigns. It has been instrumental in gathering diverse data sets, from air quality measurements to pedestrian traffic patterns.

MODE: Focusing on urban mobility, MODE gathers data on travel patterns and modes of transportation. This tool provides insights into the environmental impact of urban mobility, aiding in the development of

² The name of this tool has changed since the beginning of the GREENGAGE project. Previously used term was ‘MindView app’.

sustainable transportation policies. It has been integrated within the GREENGAGE app. Both on the server side and the client side.

Off-the-shelf environmental sensors: This project has used Atmotube PRO, which is a portable, off-the-shelf environmental monitoring device designed for real-time air quality assessment. It measures a wide range of pollutants, including volatile organic compounds (VOCs) and particulate matter (PM1, PM2.5, and PM10), as well as essential atmospheric parameters such as temperature, humidity, and barometric pressure. Compact and wearable, the Atmotube PRO offers continuous, location-aware environmental data collection, making it highly suitable for personal exposure studies and mobile environmental sensing. Its Bluetooth connectivity and integration with mobile apps allow users to access and transmit data instantly, aligning well with IoT-based environmental monitoring systems. As such, it contributes with granular, real-time data to broader environmental analysis frameworks.

The success of this area of concern hinges on active citizen participation. The tools provided are designed to be accessible and engaging, encouraging widespread community involvement in data collection efforts.

Area of Concern 3: Analysis and Visualization for Insights Generation

In the final area of concern, *Analysis and Visualization for Insights Generation*, the collected data is transformed into actionable insights. This area is where the data becomes a powerful tool for understanding and influencing environmental policy.

Apache NiFi: Serving as the gateway for data integration, NiFi facilitates the extraction and initial processing of collected data. This tool will be key in ensuring that data flows smoothly from collection points to storage and analysis systems.

Apache Druid: As a data storage solution, Druid provides a secure and efficient repository for the vast amounts of data collected. It is optimised for high-speed data retrieval, ensuring that data is readily available for analysis.

Superset, Data Quality and Structure Dashboard and UrbanTEP/VISAT: These tools are employed to create advanced visual representations of the data. Superset provides capabilities for detailed data analysis and visualization, while UrbanTEP/VISAT offers specialized tools for urban environmental data interpretation of complex datasets. These visualizations do not only aid in understanding the data but also in communicating findings to a broader audience.

DataHub: The centralization of data in DataHub ensures that all users have easy access to the information they need. This tool is instrumental in democratizing data access, allowing a wide range of stakeholders to participate in data analysis and decision-making processes, since its dashboard provides a common entry point to the data and analysis enacted in diverse Pilots' Thematic Co-Explorations.

DigitalTwin: Additionally, the DigitalTwin tool plays, within the North Brabant Pilot, a crucial role in this area. It enables the creation of a dynamic, virtual model of urban environments, integrating various data sets from previous areas of concern. This tool allows users to simulate and analyze different scenarios, providing a deeper understanding of the environmental impact of various urban policies and actions.

Continuous Iterative User Journey Flow:

The user journey in the GREENGAGE project is characterised by a continuous flow of activities, with constant interactions between the three areas and comprising several iterations. A brief depiction of this journey looks like the following:

- Users start by engaging in community building and project planning in *Area of Concern 1*, where they also continuously return for feedback and direction.
- They then proceed to actively collect data in *Area of Concern 2*, using the tools provided to capture a wide array of environmental information.
- As users gather data, they frequently revisit *Area of Concern 1* to discuss their findings, gain insights, and adjust their data collection strategies based on community input. This iterative process ensures that data collection is responsive to the evolving needs and insights of the community.
- Subsequently, in *Area of Concern 3*, users analyze and visualize the collected and curated data in *Area of Concern 2*, gaining actionable insights. These insights are again shared with the

community in *Area of Concern 1*, closing the Citizen Science (CS) loop and often leading to new questions or areas of exploration, thus initiating another cycle of data collection and analysis.

Through this interconnected and cyclical journey, depicted in Figure 2, users of the GREENGAGE project assets are not just collectors of data or recipients of information. Instead, they are active participants in a continuous process of learning, understanding, and sharing, they take part in the whole underlying Citizen Science Loop. The seamless interaction among the tools across different areas ensures that the project remains dynamic, responsive, and community-driven, truly embodying the spirit of collaborative environmental observation and sustainable urban development.

The tools selected for the areas of concern have been extracted from the Pilots' needs set out in deliverable D2.7, which is an update of D2.6. For each of the tools a general description will be given to the Observers to focus on their role in the area in which they are assigned and on the interoperability between tools. For more information on the tools see the D4.1, D4.2 and D4.9 deliverables.

Following is an in-depth description of the Citizen Observatory Community Journey depicted Figure 2 below.

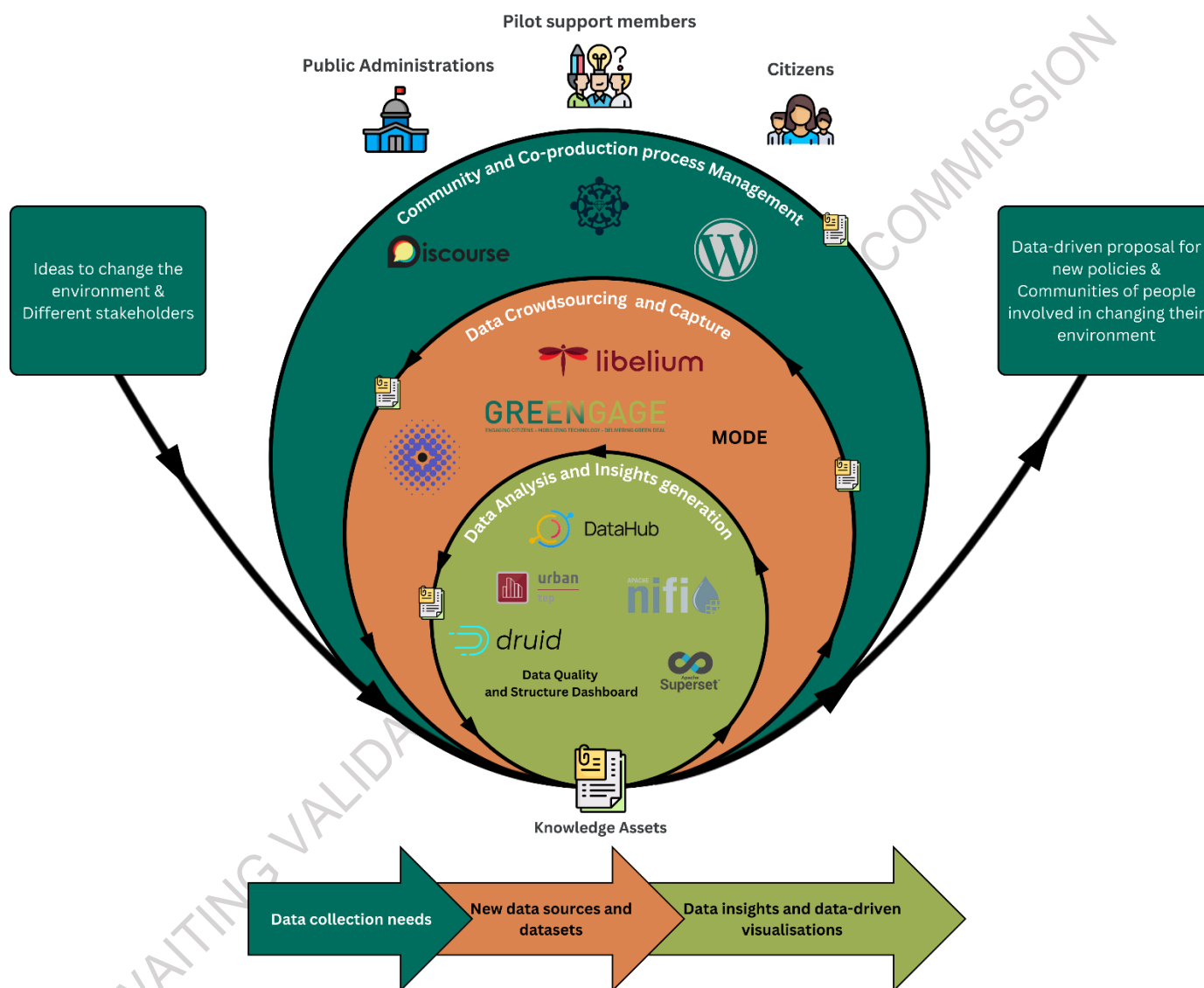


Figure 2: Overview of the Citizen Observatory Community Journey

3.1 Area of Concern 1: Community and Co-production Process Management

Area of Concern 1 of the GREENGAGE Citizen Observatory Community Journey, titled "*Community and Co-production Process Management*" is foundational to the project's collaborative and participatory framework. This area of concern emphasizes establishing and nurturing a community-focused environment, pivotal for the journey's subsequent areas. It involves setting up a digital collaborative platform where participants can initiate, develop, and maintain environmental campaigns. This platform provides tools for planning, resource allocation, and managing the campaign's lifecycle, aiming to foster a user-friendly yet robust ecosystem for information exchange and community building.

Participants engage through tools like Discourse, a centralized forum for community interaction, idea exchange, and problem-solving. In this area of concern, the role of Collaborative Environment is crucial, where stakeholders design, engage, evolve and sustain the different Thematic Co-Explorations. The focus is on transparency and ongoing dialogue, fostering a strong community sense of purpose and cohesion. Each Pilot project has its own WordPress site (<https://www.greengage-project.eu/greengage-observatories/>) to showcase progress, achievements, and impacts publicly. These integrated tools are designed for information dissemination and community empowerment, ensuring campaigns are driven by informed, collaborative efforts. This synergy and active community participation lay the groundwork for successful co-production processes, leading to effective data crowdsourcing, analysis, and policy influence in subsequent areas.

3.1.1 Collaborative Environment

GREENGAGE's Collaborative Environment (CE) is one of the results of the INTERLINK³ project and serves as a foundational tool for Citizen Observatories. Integrating the Collaborative Environment as part of the GREEN Engine will significantly enhance team communication, document collaboration, and stakeholder engagement. Its capabilities in process planning and promoting loyalty through incentives and rewards make it an invaluable asset for co-production in diverse fields. Indeed, the Collaborative Environment serves as a comprehensive tool in fostering Citizen Science and forming Citizen Observatories, facilitating collaboration, resource management, and community engagement, and creating a seamless and efficient ecosystem for citizen-led initiatives and projects.

In essence, the CE offers the following core functionalities:

- a) co-producer team and process management.
- b) guide for co-production process by means of schemas of steps to follow.
- c) recommendation of knowledge and software co-production enablers (called CO Enablers in GREENGAGE, formerly INTERLINKERS in INTERLINK project), most suitable to the problem domain represented by the chosen co-production task.
- d) selection, instantiation, assignment, claim and registry of use of CO Enablers (e.g., a Thematic co-exploration specification template or a CS campaign specification template) and
- e) CO Enablers catalogue to browse, search and review publicly available Thematic Co-Explorations' co-production enablers and processes. Access to the full catalogue of CO enablers is available at <https://demo.greengage-project.eu/catal>.

³ <https://interlink-project.eu/>

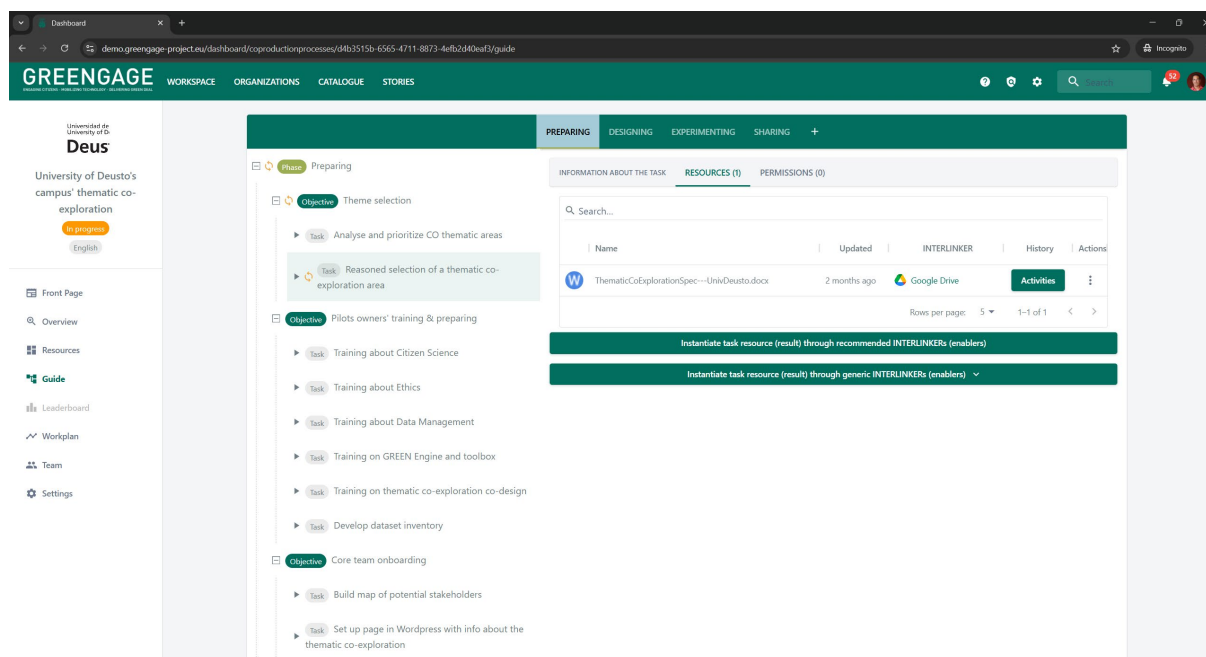


Figure 3: Example of instantiation of CO Enablers (Citizen Observatory's Thematic Co-Exploration specification template).

Role within the Citizen Observatory

A typical use case for the Collaborative Environment in GREENGAGE involves setting up and managing a co-production process within a Citizen Observatory to holistically manage the co-production, including co-design and co-delivery, of a given Thematic Co-Exploration. Users can create organizations, form teams, and develop processes, each with clearly defined phases, objectives, and tasks. The platform allows for resource management, tracking progress, and engaging communities effectively within a Thematic Co-Exploration. It also supports integration with external tools, including the GREENGAGE toolbox components, enhancing its utility in diverse scenarios.

The CE clearly governs the “Community and Co-production Process Management” area of concern. Still, it allows to plan the tasks and trace the results (called resources within the CE) that result from co-producer teams participating in the “Data Crowdsourcing and Curation” and “Analysis and Visualization for Insights Generation” areas. Indeed, the CE tool fosters Citizen Science (CS) and contributes to forming Citizen Observatories (COs) through several mechanisms:

1. **Facilitating Co-production Processes:** The environment serves as a platform where citizens (participants in a Citizen Observatory) can actively participate in co-production processes, which is a cornerstone of Citizen Science. It encourages citizens to be directly involved in collaborative processes or Citizen Science projects that matter to them, fostering a sense of ownership and engagement. Collaborative processes managed through the CE are instantiated from a set of customizable process models or schemas specially designed by GREENGAGE to offer a guideline of the different phases and tasks the co-design and co-delivery of a Thematic Co-Exploration should go through. Such schemas are a reference, do not need to be followed strictly and can be tuned to the purpose of each specific Thematic Co-Exploration. Notably, in the second iteration of the project, a new metamodel or schema to co-produce Thematic Co-Explorations in Citizen Observatories has been produced⁴. Such co-production process schema or model is documented at [link](#).

⁴ A schema to instantiate co-production processes that follow the GREENGAGE thematic co-exploration 4 phases are detailed at <https://greengage-project.github.io/Documentation/HOWTO%20Thematic%20co-explorations/>

2. **Resource Management and Sharing:** By offering a centralized system for the management and sharing of resources created during the co-production process, it promotes transparency and accessibility. This feature is vital in Citizen Observatories where various resources like training materials, communication materials, data sets and research findings (e.g., visualization or policy briefs) can be shared and managed efficiently, promoting collaborative research and knowledge sharing.
3. **Task Management and Tracking:** The functionalities to manage phases, objectives and tasks and track the progress of various activities within a co-production process help in organizing and managing citizen-led initiatives. This ensures that tasks are completed efficiently, and objectives are met, which is crucial for the success of Citizen Science projects. Although, there can be a guideline or process schema to which many Citizen Observatory co-production processes may comply, each specific Citizen Observatory may adapt such process, removing tasks, adding new tasks, or renaming them, among other things.
4. **Community Engagement and Collaboration:** The environment promotes community engagement by allowing the creation of organizations and teams. It fosters collaboration among citizens, public administrations, non-profit organizations, and other stakeholders, facilitating a cooperative approach to problem-solving and project implementation, which is a fundamental aspect of Citizen Observatories.
5. **Integration with External Tools:** The ability to integrate with external tools and assets, referred to as "INTERLINKERS" in the original project this tool was designed for, namely, INTERLINK, or CO Enablers using REENGAGE jargon, facilitates a seamless flow of information and functionalities between different platforms. This integration is essential in forming a cohesive Citizen Observatory where various tools and resources can be linked and combined effectively. Such tools are exposed as part of the CE catalogue and REENGAGE Academy, this latter one also accessible from the project's WordPress based website.
6. **Replicating Success Cases:** The tool offers the functionality to clone successful co-production processes, serving as a repository of best practices and good successful past co-produced processes. This can be combined with other tools to replicate success stories in different contexts, promoting knowledge sharing and learning, which is vital in expanding and enhancing Citizen Observatories.
7. **Gamification Engine:** The platform offers a gamification/incentivization engine that provides points to users depending on their contribution to different assets. The gamification engine and the points each task provides are activated by the administrators/managers of the process. The platform also provides a leader board to compare your contribution with other participants.

Interoperability with other tools

The Collaborative Environment may be hyperlinked from REENGAGE's project website and the pages dedicated to each Pilot site and their associated Thematic Co-Explorations. Per Thematic Co-Exploration a new co-production process shall be instantiated and hyperlinked from the WordPress site. Likewise, each new Thematic Co-Exploration's co-production process may be associated to a brand-new topic in the Discourse tool where communication and dialogue among fellow citizen observers and related stakeholders could mainly take place. On the opposite direction, the instantiated co-production process per Thematic Co-Exploration will also include as new resources: a) one external resource, i.e. link pointing to the page in WordPress where the Thematic Co-Exploration is described, and b) another external resource pointing to the Discourse topic where discussions on this Thematic Co-Exploration are being carried out.

The CE can internally instantiate generic resources common in any type of co-production process, e.g., documents, spreadsheets, presentations, questionnaires, augmented webpages by Augmenter internal tool or decision-making spaces made available through Loomio⁵. Besides, it can link to "External resources" which are produced through external tools not internally integrated within the CE, e.g., a workflow defined in Apache Ni-Fi, a large CSV file stored in Apache Druid, a visualization created through Apache Superset. If the external tool used to produce a resource enables the generation of permalinks, these can be easily cross-linked from the corresponding task in the co-production process governing a

⁵ <https://www.loomio.com/>

Thematic Co-Exploration. For example, when the citizen observers leading a Thematic Co-Exploration want to carry out the objective “Data combination, analysis & visualization” (see Figure 4) defined in the reference GREENGAGE co-production process schema, they would associate to such task new external resources pointing to each of the workflows used to aggregate, mix and match heterogenous datasets generated by Apache NiFi and they would also include an external resource link or reference for each visualization generated over the aggregated datasets produced with the support of Apache Superset tool. Hence, data integration can be governed from within the CE and its results traced as external resource entries within the Thematic Co-Exploration processes’ tasks. The actual data mining and analysis process are carried out by external tools configuring the GREENGAGE toolbox. However, their usage and results tracking (or provenance) can be performed by the CE.

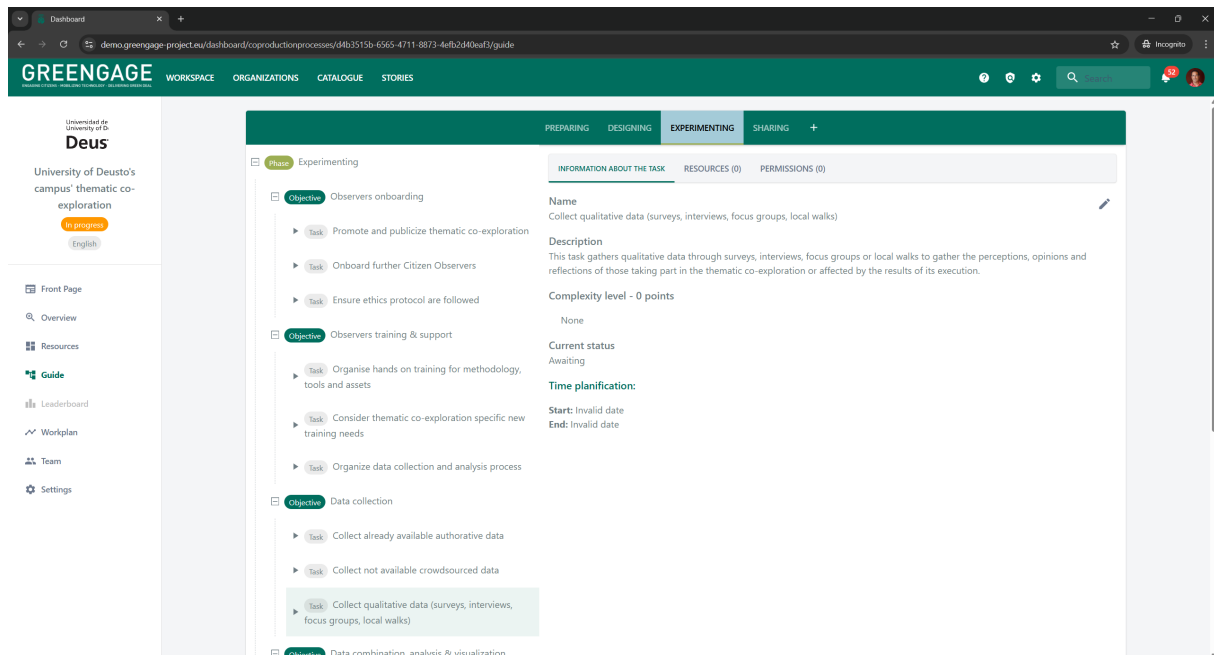


Figure 4: View in Collaborative Environment of the reference co-creation process model for Thematic Co-Explorations.

The user experience continuity on a given Thematic Co-Exploration is guaranteed by CE since it allows governing the underlying co-production process and the citizen observers’ teams and related stakeholder teams.

- Participants in a task of a co-production process can assign tasks to other citizen observers.
- Participants can also claim having contributed to co-produce certain resources, e.g., an Apache NiFi workflow and be rewarded for it, once the Thematic Co-Exploration moderators validate their contributions.
- Newly created resources in the form of visualizations generated with Apache Superset, as results of collaboration in each task, e.g., “Report the results”, can be shared with participants in the co-production process or even external users. The environment only allows permalinks to be connected as results from the co-production process tasks.
- Needs for new observers to participate task, e.g., participation on air quality data crowdsourcing campaign, can be declared by making a process public and, hence, enable new citizens to sign in for collaboration.
- When a Thematic Co-Exploration is considered as finished or at least to have produced a Minimal Viable Product (MVP), the underlying co-production process, its resources and different metadata such objectives, results, owners, licenses and so on, can be published as success stories.

Visit the Collaborative Environment (CE) under <https://demo.greengage-project.eu/>. This link will guide you to the CE's user interface, where you can interact with the various tools and functionalities it offers.

The process model created for governing the co-creation process associated to Thematic Co-Explorations is depicted in Figure 5.

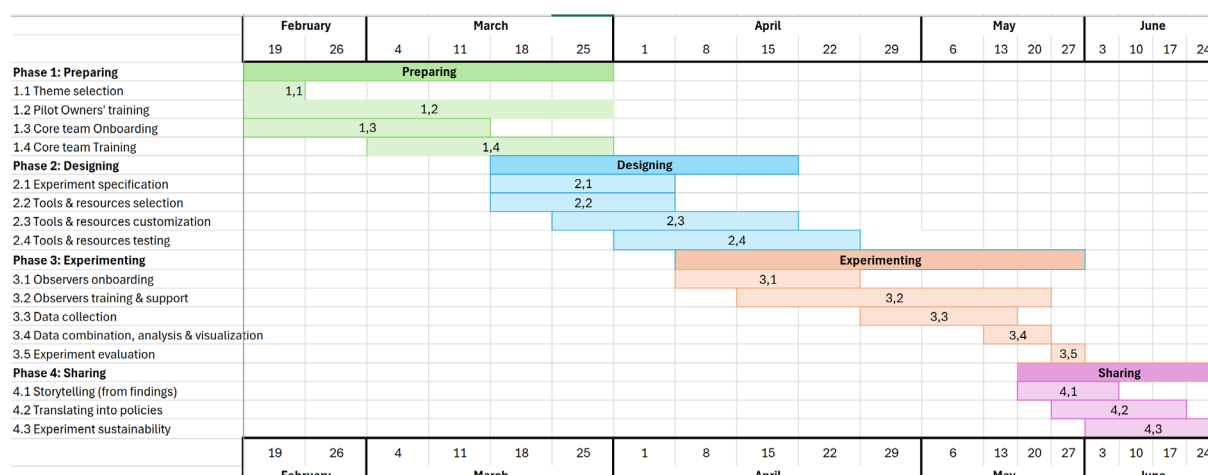


Figure 5: Reference process model created to govern Thematic Co-Explorations' co-creation.

3.1.2 Discourse

Discourse is an open-source discussion platform tailored for building a community forum, ideal for projects like Citizen Observatories. Its comprehensive feature set focuses on enhancing user engagement and streamlining conversations. The platform is designed to transform the way forum discussions are conducted. Discourse offers a continuous, just-in-time loading conversation stream, ensuring the dialogue flows smoothly without any interruptions. This dynamic approach to conversations makes it easier for users to follow and participate in discussions. Additionally, Discourse integrates real-time chat channels, fostering informal, immediate interactions within the community. These chat messages can seamlessly transition into more structured forum topics, maintaining both continuity and discoverability of discussions.

Discourse stands out through its integration of Artificial Intelligence, which significantly improves community interaction and moderation. This makes the platform not only more user-friendly but also more efficient in managing community dynamics. Among its unique features, Discourse offers extensive personalisation options. Users can adjust their experience through a custom sidebar and user preferences, while admins can set defaults and site settings for a configurable community environment. The platform also maintains a simple, linear conversation flow, with contextual expansion in posts and quotes to reveal full conversations without disrupting the user's navigation. Moreover, sharing links on Discourse is interactive and informative, as pasted links automatically expand to show additional content and context from various popular websites.

From a technical standpoint, Discourse boasts several impressive specifications. It supports Single Sign-On (SSO) integration with Keycloak⁶ using OpenID⁷ Connect, ensuring user convenience and security. Additionally, the platform includes community moderation tools for managing spam and resolving conflicts, fostering a healthy forum environment. It also features built-in protections like Akismet⁸ spam filtering and new user restrictions to block spam effectively. Email integration allows members to stay connected by receiving notifications and replying directly via email, ensuring they remain engaged even when offline. The extensive plugin ecosystem and comprehensive API provide additional functionalities and customisability. As a 100% open-source platform, Discourse encourages community-driven development and inspires confidence. Lastly, its one-click upgrade feature simplifies maintenance, making updates hassle-free.

⁶ <https://www.keycloak.org/>

⁷ <https://openid.net/>

⁸ <https://akismet.com/>

Role within the Citizen Observatory

Within the framework of a Citizen Observatory, Discourse can serve as a very valuable engagement and communication platform. The platform can foster an environment where community members can actively participate in sharing observations, discussing findings, and collaborating on Citizen Science projects. It can also serve as “safe space” for communication amongst the Observers. This environment is conducive to an exchange of ideas and experiences, enhancing the collective understanding and co-production processes essential in Citizen Science initiatives.

The utility of Discourse extends to various facets of community and co-production process management. It stands as a hub for generating discussions, enabling participants to debate ideas, share experiences, and engage in collaborative efforts. In the realm of data crowdsourcing and curation, Discourse provides a venue where users cannot only share the data they have collected but also discuss their implications, contributing to a collective effort of refining and curating this information. This is also a shared functionality with the Collaborative Environment. Furthermore, while Discourse may not be primarily a data analysis tool, it plays a significant role in facilitating discussions around data interpretation and visualization. Insights gleaned from these discussions can then be visualized or analyzed further using other tools. Beyond its data-centric capabilities, the platform is particularly adept at fostering community engagement, facilitating Q&A sessions, and encouraging collaborative problem-solving, all of which are vital for the successful operation of a Citizen Observatory.

To improve the usability of the Discourse platform and strengthen participant engagement within the Citizen Observatories, several targeted plugins have been deployed, during the second iteration of the solution. These tools have been selected and configured to address specific functional needs and to support more effective interaction among users within the REENGAGE project framework.

The **Automation**⁹ plugin permits administrators to execute scripted actions based on defined triggers. In conjunction with the **Custom Wizard**¹⁰ plugin, it has been used to create a landing wizard (see Figure 6) that guides new users through a process of initial customisation. This includes the identification of the Citizen Observatory with which the user is associated and the username they want to display at Discourse.

⁹ <https://meta.discourse.org/t/discourse-automation/195773>

¹⁰ <https://meta.discourse.org/t/73345>

Figure 6: Landing wizard for new users.

The **Category Experts**¹¹ plugin fosters a community-based endorsement system, wherein users can recognise each other as experts within particular thematic categories. This functionality supports the visibility and authority of knowledgeable participants within the Citizen Observatories, thereby promoting peer-to-peer learning and recognition.

To encourage active participation, the **Gamification**¹² plugin enables the configuration of community contests that reward user engagement. Administrators can design scoring systems and display leaderboards to highlight individual contributions and foster a sense of progression and achievement.

The **Solved**¹³ plugin introduces a mechanism for designating accepted answers within discussion threads in specified categories. This feature enhances the clarity and navigability of the discourse by making validated solutions more accessible to other users.

The **Translator**¹⁴ plugin, integrated using the open-source tool *LibreTranslate*¹⁵, facilitates multilingual interaction on the platform (see Figure 7 and Figure 8). By enabling users to translate posts directly within Discourse, this tool supports communication among participants from different linguistic backgrounds, ensuring that language differences do not constitute a barrier to collaboration or information exchange within the GREENGAGE consortium.

¹¹ <https://meta.discourse.org/t/190814>

¹² <https://meta.discourse.org/t/225916>

¹³ <https://meta.discourse.org/t/30155>

¹⁴ <https://meta.discourse.org/t/32630>

¹⁵ <https://libretranslate.com/>

How to Participate

- Join discussions in the relevant subcategories based on your interests.
- Share your thoughts, experiences, and ideas from the pilot event.
- Stay connected with other participants, experts, and local stakeholders.
- Contribute to shaping future initiatives based on the pilot's findings.

Get Started

👉 **Introduce yourself** in the comments below! Tell us which group you were part of and what excites you most about the GREENGAGE project.

👉 **Visit the subcategories** Attractiveness & Livability group and Mobility group to dive into focused discussions.

Let's keep the momentum going and work together toward a greener, smarter Turano Valley! 🌱



🌐 ❤️ 🔗 ... ➡ Rispondi

Figure 7: A Discourse topic in which the automatic translation button can be seen.

Get Started

👉 **Introduce yourself** in the comments below! Tell us which group you were part of and what excites you most about the GREENGAGE project.

👉 **Visit the subcategories** Attractiveness & Livability group and Mobility group to dive into focused discussions.

Let's keep the momentum going and work together toward a greener, smarter Turano Valley! 🌱

🌱

Chi siamo Turano Valley Pilot (Marzo-2025) categoria

Tradotto dalla lingua en da LibreTranslate

Benvenuti nello spazio di discussione per il **Turano Valley Pilot** della **GREENGAGE** progetto! Questa categoria è dedicata a continuare la conversazione al di là del nostro evento in loco, permettendo ai partecipanti di rimanere impegnati, condividere intuizioni, e collaborare nella formazione di una valle Turano più sostenibile e vibrante.

Informazioni sul pilota

Il Turano Valley Pilot esplora soluzioni innovative per migliorare **sostenibilità, mobilità e vivibilità** nella regione. È strutturato intorno a due gruppi di lavoro chiave:

- **Attrattiva e Livibilità** – Concentrati sul rendere Turano Valley un luogo più invitante e fiorente per residenti e visitatori.
- **Mobilità** – Esplorare i miglioramenti per aumentare la sicurezza stradale sulla Valey.

20 mar
1 / 7
20 mar
13g

Figure 8: Result of the automatic translation of a Discourse topic.

Interoperability with other tools

Discourse's advanced interoperability features, particularly its robust API, facilitate seamless data sharing and integration with various tools. The API enables Discourse to effectively receive and incorporate data from diverse sources such as WordPress posts and various data analysis platforms, allowing this information to be displayed and discussed within forum threads. This integration capability is detailed in the Discourse API Documentation, represented in Figure 9. Additionally, the platform supports embedding and linking functionalities, which allows users to effortlessly embed or link visualizations and data from other platforms directly into their posts. This level of integration ensures that Discourse acts not just as a standalone forum but as a comprehensive hub for data interaction and collaborative discussion.

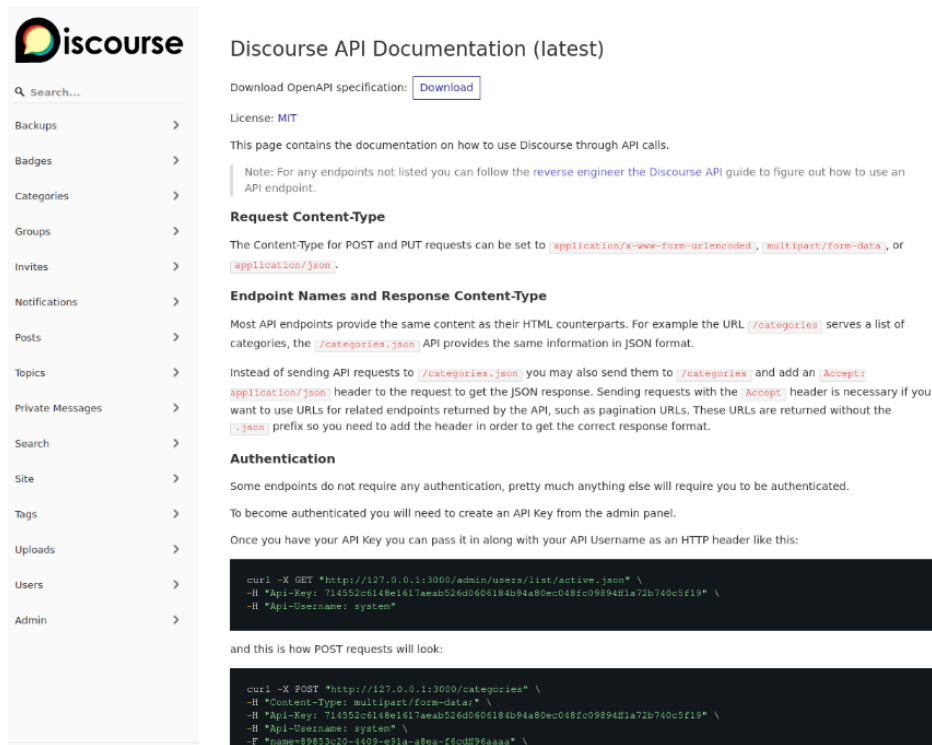


Figure 9: Discourse API Documentation.

The seamless transition between phases in a project is a key aspect of the user experience in Discourse. Insights and discussions on the platform can significantly inform and shape activities in other stages of the project. For instance, a lively discussion about a specific data trend on Discourse can lead to a more focused approach in data collection or analysis in subsequent areas. Moreover, the platform facilitates cross-tool engagement. Users can reference discussions, polls, or insights gathered on Discourse in other tools, ensuring a consistent and uninterrupted thread of engagement and knowledge sharing throughout the entire lifecycle of the project. This feature of Discourse greatly enhances the coherence and continuity of the user experience, making it an asset in any collaborative endeavour. Notice that Discourse and Collaborative Environment share some features for sharing contents and fostering dialogue. They should be understood as complementary and could be used simultaneously or independently in different Thematic Co-explorations. Whilst Discourse is designed to foster community engagement and dialogue, the focus on the Collaborative Environment is more on the management and governance of co-production processes. Whilst Discourse is more accessible and could be adopted by every observer in a Thematic Co-exploration, the Collaborative Environment is a more complex and sophisticated tool, which is more targeted to more active well-trained volunteers, which are responsible for completing the different stages of a Thematic Co-exploration.

3.1.3 WordPress

WordPress is a versatile Content Management System (CMS). It initially gained popularity for blog creation but has since evolved into a key tool for developing a wide range of websites, including commercial ones. Its flexibility and user-friendly interface make it a go-to solution for website creation. Its primary objective is to simplify the process of website creation and management. Its user-friendly design, combined with powerful features, makes it an ideal choice for users ranging from novices to seasoned web developers. WordPress aims to enhance website functionality, improve user experience, and provide robust website management tools. Its main features include:

- Easy Installation, Update, and Customisation: Streamlines the setup and maintenance process.
- Automatic System Updates: Ensures the CMS stays current and secure.
- Multiple Authors and User Roles: Allows for collaborative work with various permission levels.

- **Support for Multiple Blogs:** Facilitates the management of several blogs within a single site.
- **Static Page Creation Capability:** Enhances the flexibility of content management.
- **Additional Features:** Includes categorization of articles and pages, password protection, WYSIWYG editor, email publishing, import options from various platforms, auto-save drafts, comment and communication tools, permalink support, content and comment distribution via RSS and Atom, link management, multimedia file management, plugin and widget support, integrated search functionality, and theme creation system.

Role within the GREENGAGE project

Within the GREENGAGE project, WordPress serves as the platform for the development and management of the GREENGAGE website (www.greengage-project.eu). Its versatility and extensibility via plugins, makes it the ideal choice to meet the project's objectives of accessibility, user engagement, and content dissemination.

The website is formed by the following main sections:

- **Become a GREENGAGE Observer:** In this section, visitors are invited to actively participate in the project by joining one of the GREENGAGE Observatories. It explains how citizens, local authorities, and organisations can contribute to data collection and environmental monitoring in their communities. The section encourages collaborative engagement and outlines the benefits of becoming an observer, such as contributing to local policymaking and accessing tools for environmental action.
- **GREENGAGE Academy:** This is the project's educational and training hub. It provides open-access learning materials, toolkits, guidelines, and resources to support citizens, practitioners, and policymakers involved in the observatories. The Academy fosters capacity-building by promoting environmental literacy, digital skills, and participatory governance methods.
- **GREEN Engine Platform:** This section introduces the **digital backbone of the GREENGAGE project**—a comprehensive ecosystem that enables and supports the thematic co-production processes within the GREENGAGE Observatories. The GREEN Engine provides an integrated suite of digital tools, online services, and portable devices designed to empower citizen-led environmental exploration and action.
- It guides participants through the entire data lifecycle: from **data collection and validation** using mobile apps and IoT sensors, to **analysis and visualization** through advanced platforms, ensuring that local knowledge can be transformed into actionable environmental insights.
- The platform's **digital ecosystem** is organised into three interconnected domains:
 - **Community and Co-production Process Management:** This includes tools like the *Collaborative Environment (CE)* for planning and managing Citizen Science campaigns, *Discourse* for forum-based community dialogue, and *WordPress* pages dedicated to each pilot site for public engagement and outreach.
 - **Data Crowdsourcing and Curation:** Tools such as the *MindEarth for GREENGAGE* app, *GREENGAGE* app, *MODE* mobility data tool, and *IoT hyperlocal sensors* (e.g. *Atmotube PRO*) enable real-time data collection on environmental conditions, urban mobility, and air quality—all driven by citizen participation.
- **Analysis and Visualization for Insights Generation:** The GREEN Engine uses a robust data infrastructure including *Apache NiFi* for data integration, *Apache Druid* for storage, and *Apache Superset* for interactive dashboards. *VISAT/UrbanTEP*, the *Data Quality Dashboard*, and the *DataHub*¹⁶ further enhance the project's ability to explore, analyze, and share complex environmental data.

¹⁶ <https://datahub.greengage-project.eu/>

Altogether, the GREEN Engine equips GREENGAGE Observatory participants with the necessary digital tools to co-produce knowledge, promote data transparency, and support informed decision-making in pursuit of more sustainable and inclusive urban environments.

- **GREENGAGE Observatories:** Here, users can explore the GREENGAGE Observatories within the five GREENGAGE Pilots (Bristol, Copenhagen, Gerace, Turano Valley, and North Brabant). Observatory pages provide local context, objectives, activities, and ways in which citizens and stakeholders work together to address urban environmental challenges.

Knowledge Base: A collection of deliverables, datasets and publications generated during the project, made available in time of their creation.

- **Events:** This section provides information about upcoming and past events related to the GREENGAGE project. It includes workshops, conferences, field activities, and meetings, showcasing how the project fosters collaboration and knowledge exchange among partners, communities, and experts.
- **Blog:** The blog shares news, stories, and updates from the GREENGAGE project. It includes highlights from meetings, citizen initiatives, technical updates, and success stories from the observatories. This section helps keep the public and stakeholders informed and engaged throughout the project's lifecycle.
- **'What is GREENGAGE?':** Here is a comprehensive overview of the GREENGAGE project, including its goals, methodology, timeline, and the consortium of partners behind it. It also includes acknowledgments of funding sources, contact information, and links to legal documents such as privacy and cookie policies.

Interoperability with other tools

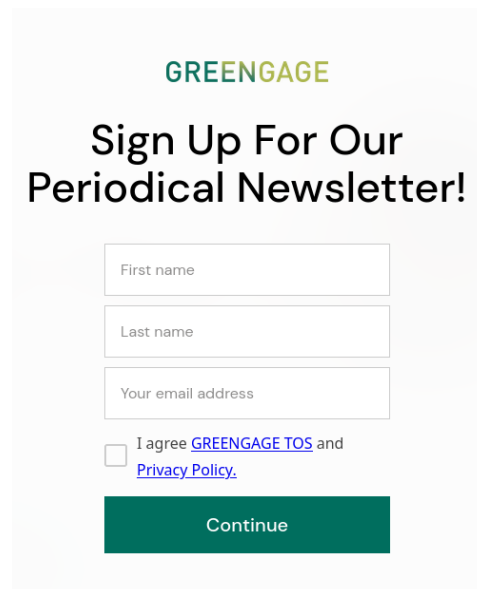
The GREENGAGE project website benefits from a diverse suite of WordPress plugins that enhance its functionality, performance, interactivity, and compliance. These integrations support the user experience, data handling, communication, and legal adherence essential for a large-scale EU-funded initiative. Below is an overview of how the installed plugins contribute to the overall integration and capabilities of the site:

- **The Events Calendar** and **Events Shortcodes for The Events Calendar** allow the publication and display of upcoming events in flexible layouts, supporting visibility for workshops, campaigns, and co-production sessions.
- **Filter Everything PRO** and **Filter for Divi** improve content navigation by allowing visitors to filter posts, datasets, or observatory pages by specific criteria, supporting better user orientation and content discovery.
- **MapGeo – Interactive Geo Maps** facilitates the integration of interactive geographical maps to showcase project locations, observatories, and campaign coverage areas, making spatial data more accessible and visually engaging.
- **Complianz | GDPR/CCPA Cookie Consent** ensures that the website complies with European and international data protection laws by managing cookie consent banners, cookie policy generation, and conditional script loading based on user consent.
- **OMGF** self-hosts Google Fonts to improve GDPR/DSGVO compliance by reducing external data requests and increasing privacy.
- **Connect Matomo** integrates the privacy-friendly Matomo analytics platform directly into the WordPress dashboard, allowing the team to monitor visitor behavior without compromising data protection principles.
- **W3 Total Cache** significantly enhances site speed and performance through caching mechanisms, content delivery network (CDN) integration, and minification of assets, all of which contribute to a smoother user experience.
- **UpdraftPlus** provides automated and secure backup solutions, safeguarding the website's data and configurations against unexpected failures or cyber threats.

- **Query Monitor** offers developer-level diagnostics and debugging tools, helping the web team to monitor performance, identify bottlenecks, and optimise plugin interactions.

These plugins collectively ensure that the GREENGAGE website not only serves as a functional and informative platform but also acts as an interactive, legally compliant, and high-performance digital hub for Citizen Science collaboration across Europe.

In addition, GREENGAGE website and Mailchimp have been integrated to allow visitors to easily subscribe to the GREENGAGE newsletter. As shown in Figure 10, a pop-up form is shown to all new visitors.



The image shows a pop-up form for signing up for the GREENGAGE newsletter. At the top is the GREENGAGE logo. Below it is the heading "Sign Up For Our Periodical Newsletter!". The form contains three input fields: "First name", "Last name", and "Your email address". Below these fields is a checkbox with the text "I agree [GREENGAGE TOS](#) and [Privacy Policy](#).". At the bottom of the form is a green button labeled "Continue".

Figure 10: Mailchimp newsletter subscribing form.

3.2 Area of Concern 2: Data Crowdsourcing and Curation

Area of Concern 2 of the GREENGAGE project, titled "*Data Crowdsourcing and Curation*" marks a pivotal stage in the Citizen Observatory Community Journey. This area is characterized by the active engagement of participants in gathering essential environmental data, leveraging state-of-the-art technology. It involves users employing various tools like MindEarth for GREENGAGE app, GREENGAGE app, MODE, and IoT of-the-shelf sensors to collect and curate data from their surroundings. These tools enable citizens to contribute meaningfully by collecting data on aspects such as air quality, urban mobility, and environmental conditions.

This area is designed to be interactive and user-friendly, encouraging broad community participation in data collection efforts. The success of this area of concern hinges on the active and enthusiastic involvement of citizen observers, who play a direct role in building a comprehensive environmental database through MindEarth for GREENGAGE app and GREENGAGE app. This collaborative effort not only enriches the data pool but also empowers citizens by giving them an active role in environmental monitoring and analysis. The data collected during this area of concern is crucial for later stages of a Thematic Co-Exploration process, where the data is analyzed and visualized to generate insightful findings and inform environmental policies.

3.2.1 MindEarth for GREENGAGE app

The MindEarth for GREENGAGE app (see Figure 11), available on Google Play and as a standalone Android Package Kit (APK) for partners, is a sophisticated tool designed for geospatial data collection. The app is also available for iOS users through Apple TestFlight, extending accessibility and allowing the platform's activities to reach a wider range of contributors across devices. It enables users, regardless of their technical background, to capture and upload geo-tagged street-level imagery using

smartphones or compatible 360° cameras. This tool is integral for citizen-powered geospatial data collection, facilitating the acquisition of images of streets, landscapes, landmarks, and other visible geographic features in public spaces. These images are then processed through advanced AI algorithms to create and enhance detailed maps.

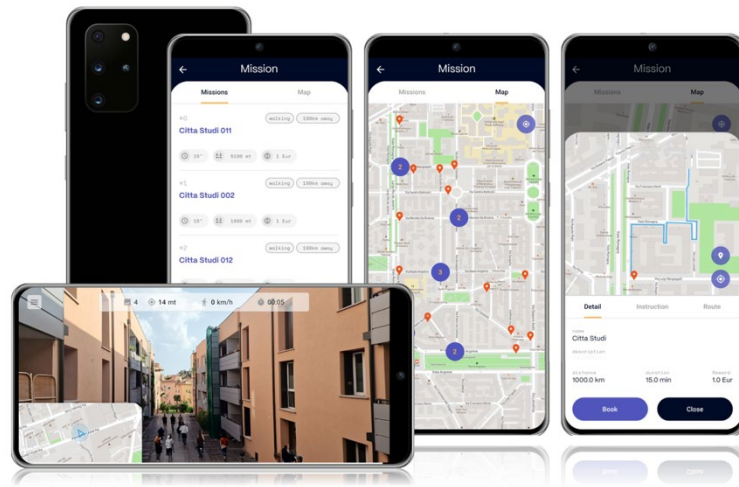


Figure 11: MindEarth for GREENGAGE app running on smartphones.

After registering, users become "mappers" and opt to undertake various "missions" associated with specific physical locations, detailed in the Mission Brief. These missions encompass a range of tasks, including visiting certain streets or areas, with specific start and end points and paths to follow. Each mission includes a title, description, estimated distance and time for completion, and a reward for successful completion. The MindEarth for GREENGAGE app provides notifications for new missions in the user's vicinity, guiding them with tutorials and highlighting the specific requirements of each mission.

Missions are created and uploaded through MindEarth's "Mission Control" platform, which can be operated by MindEarth or authorised third parties. This involves generating the specific geometric shapes and time constraints for each mission in a GeoJSON file and a specification file. In addition, the platform allows authorised users to manually draw circuits or paths directly within the Mission Control interface, offering greater flexibility and autonomy in mission design, which is particularly beneficial for Pilot-specific implementations.

In compliance with GDPR regulations, the system anonymises images containing personal information through advanced machine learning algorithms. Upon collection, images are initially stored with encryption on the local drive of the mapper's device while full image anonymisation (i.e. licence plates and faces) occurs after being uploaded in a secure AWS environment on servers located in EU. At this point un-anonymised imagery are permanently deleted.

The anonymised images are then used to generate aggregated data layers. These layers include detailed insights into urban environments and social dynamics, such as:

- **Housing and Infrastructure Data:** Condition of housing, roads, bridges, public amenities, and utilities.
- **Green Space Analysis:** Characterisation and maintenance status of green spaces and ecological features.
- **Socio-Economic and Cultural Insights:** Information derived from visual cues like commercial activities, human presence, public gatherings, and other social dynamics in public spaces.
- **Mobility and Traffic Patterns:** Data on the presence of different types of vehicles, pedestrian flows, and overall mobility within urban areas.

This processed data is crucial for various analytical purposes, including urban planning, foot traffic analysis, public space utilisation, and disaster risk management. The aggregated data layers are managed in the MindEarth cloud backend, ensuring transparency for the end-user and secure handling of sensitive information. This comprehensive approach enables MindEarth for GREENGAGE app to play

a pivotal role in urban and spatial planning initiatives, leveraging citizen-contributed data to enhance public understanding and management of urban spaces.

Role within the Citizen Observatory

The app plays an integral role in the "Data Crowdsourcing and Curation" area of concern. It addresses a key Citizen Science challenge in geospatial data collection and mapping, enabling users to actively contribute to the creation and improvement of comprehensive maps, in a way that requires no pre-existing skills or knowledge. This is achieved thanks to the app's user-friendly interface and engaging game-like experience, which simplifies the process of capturing and uploading geo-tagged images of streets, landscapes, landmarks, and other geographical features using smartphones or compatible devices.

MindEarth for GREENGAGE app's mission-based approach is central to its functionality. Users, even those with limited technical experience, can undertake various missions, involving tasks like following predefined routes or capturing specific objects. The incentive of completing missions and receiving rewards fosters active community participation in data gathering, making it accessible and engaging for a broad audience.

This approach is particularly useful for maintaining up-to-date and comprehensive maps, a task that can be resource-intensive for official mapping agencies. This is particularly valuable in regions with limited or outdated official mapping sources. By harnessing the collective efforts of the community, especially in less documented or rapidly changing areas, MindEarth for Greengage app fills critical gaps in mapping data.

Providing real-time, on-demand photographic surveys of urban spaces, MindEarth for Greengage app can be used to assess pedestrian presence and foot traffic patterns in busy city centers. Such insights are invaluable for optimising urban planning and enhancing public space utilisation, demonstrating MindEarth's app versatility and its critical role in community-driven spatial knowledge and infrastructure development within the Citizen Observatories of GREENGAGE Pilots.

Interoperability with other tools

The MindEarth for GREENGAGE app does not rely on third-party data sources and instead gathers data directly from smartphone applications and a crowd-sourcing approach. This approach ensures the authenticity, timeliness and relevance of the data collected via street-level mapping campaigns. The data thus collected is extensively processed using Computer Vision techniques, focusing on various thematic aspects relevant to the objectives of each Pilot and Thematic Co-Exploration.

Aggregated data collected through MindEarth for GREENGAGE app is accessible to the public through web APIs developed by MindEarth, either in form of a 2D result map or as anonymised images with identified objects marked as a coloured frame. These APIs expose the database content following an agreed schema, thereby facilitating its use across diverse applications as envisioned by Pilot owners. Notably, there is no need for any installation to access and use these APIs, as a public Swagger¹⁷ interface is provided for interactive development. This ensures ease of access and flexibility for developers and researchers.

In addition to public access, a copy of all aggregated metrics and data is systematically transferred to the UWE OneDrive. This data is made accessible to all project partners in GeoJSON format or as tabular data (CSV).

Moreover, the aggregated data derived from Computer Vision analysis of people, built environment features, vehicles and green elements present in street-level images, can be visualized or analyzed in other tools. An example of such a tool is the web-based analytical platform framework provided by GISAT. This platform has been previously employed in various customer implementations and is adept at visualizing maps that incorporate MindEarth's results, contributing to the project's data-rich [framework](#).

¹⁷ <https://swagger.io/>

To enhance integration within the GREEN Engine toolkit, a specialized MindEarth for GREENGAGE app branch has been developed. Key functionalities for seamless user experience across GREEN Engine tools include:

- **Centralized User Authentication System:** The MindEarth for GREENGAGE app has adopted a centralized user authentication system provided by DEUSTO, which is shared among all tools within the GREEN Engine tools portfolio. Notably, account creation to MindEarth for GREENGAGE app is handled through a DEUSTO-managed external service, redirecting users outside the app via a predefined link.
- **Integration of MindEarth for GREENGAGE app with GREENGAGE app:** To streamline user interactions and mission management between different crowd-based data collection tools, a dedicated integration protocol between the MindEarth for GREENGAGE app and the GREENGAGE app was implemented. The implementation of a dedicated protocol ensures availability and visibility of MindEarth's missions across both apps, with seamless redirection for mission execution. This means that any mission available in MindEarth for GREENGAGE app will also be visible in the GREENGAGE app and that users of the GREENGAGE app will be able to select a MindEarth for GREENGAGE app mission and be redirected to the MindEarth for GREENGAGE app to complete the booking and mission execution and upload process, as well as to know, in real time the status, duration, reward and deadline of MindEarth for GREENGAGE app's missions. This involves using two main systems.
 - **Webhook:** A webhook functionality for actions like booking, cancelling, and completing missions. These webhook interactions use the base path <https://mindearth.greengage.dev/>.
 1. Mission Booking Interaction:

Action: Booking a Mission on MindEarth for GREENGAGE app

URL: https://mindearth.greengage.dev/mission/<mission_id>/book

Method: PUT

Header: Bearer <token>
 2. Mission Cancellation Interaction:

Action: Cancelling a Mission Reservation on MindEarth for GREENGAGE app

URL: https://mindearth.greengage.dev/mission/<mission_id>/book/cancel

Method: PUT

Header: Bearer <token>
 3. Mission Completion Interaction:

Action: Completing a Mission on MindEarth for GREENGAGE app

URL: https://mindearth.greengage.dev/mission/<mission_id>/complete

Method: PUT

Header: Bearer <token>
 4. Back-Office Integration:

Action: Adding New Missions

URL: <https://mindearth.greengage.dev/mission>

Method: POST

Header: Bearer <token>

Body: Details provided in the relevant JSON structure documentation.
 - **Deep linking** facilitates user transition between GREENGAGE and MindEarth for GREENGAGE app. Deep links in each mission enable seamless transition from the GREENGAGE app and engagement with MindEarth for GREENGAGE app. A query

URL provides a unique webhook link for each mission, eliminating the need for an Auth-Header due to the signed nature of the request. Example of Deep Link:
mindearth://mission/23?callback=https%3A%2F%2Fmindearth.greengage.dev%2Fmission%2F23%3Fsignature%3Df834ed8570e05de6c50ad10bd6abcf71e9867fcb14bdf2670b4bf572ce346f3b

3.2.2 GREENGAGE app

The GREENGAGE app is an application designed for cities to intensify citizen participation in the policy making process via collection of relevant research data and/or suggesting further improvements/raising issues on a location basis. As such, it is a central hub for a) communicating research in the form of missions to the citizens b) collecting data through engaged citizen participation c) providing data to further tools for evaluation and visualization. It consists of two components.

On one hand, the Android and iOS application (which integrates MODE - see below - among other things) and the backend, which can serve as a hub for data usage by using state-of-the-art API concepts. The advantage of both components is that they are highly modular and yet designed for shared use.

For maintenance, creation, and simplified access to data, there is a user-friendly and intuitive portal available at console-stage.greengage.dev. This portal allows users to view MODE analyses, create observatories, manage users, place point of interests, create tasks, and easily download collected data. The primary goal is to provide a platform that empowers users without requiring any technical expertise.

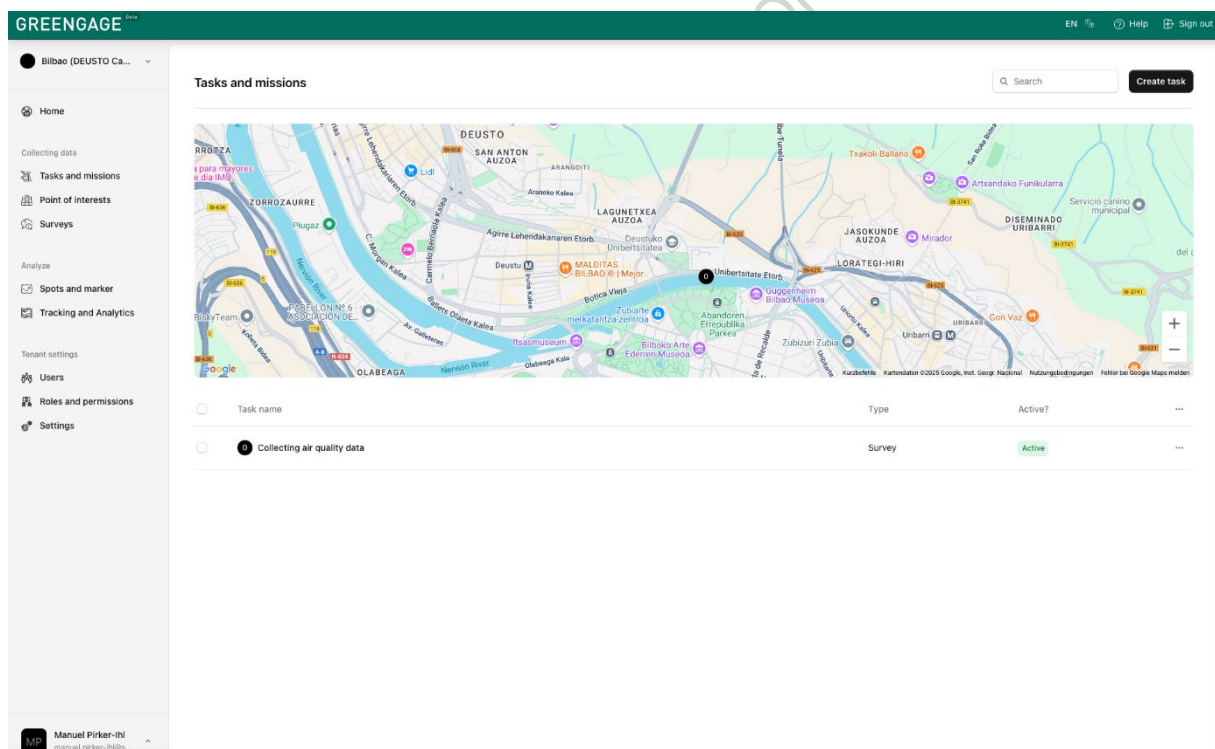


Figure 12: GraphQL view of the GREENGAGE app admin console.

Role within the Citizen Observatory:

The role of the GREENGAGE app is to make a wide range of information and technologies accessible in the simplest possible way. By using the application, participants are enabled to interact with the policymakers of their city and contribute to the research, on which the policies are further based. They can fulfil missions and access a wide variety of data sources (e.g., MindEarth) and technologies (e.g., MODE). Frictionless access for end users is the focus here.

To reduce friction, the API was simplified in the second iteration, and a streamlined Admin Control Center was introduced (Figure 12). This interface allows users to create an Observatory in just two clicks—

without the need for a developer. For developers, however, there are extended options to integrate and customise functionality via the GraphQL API.

The API part offers an easy entry into the exploration of the generated data through its use of GraphQL (Figure 13). By using the GraphQL Federation Standard, additional interfaces can also be made accessible in this way.

Interactive part of GREENGAGE app itself is used to aggregate, validate and interlink data. Tasks can be defined by the admins (and users in the later stage – optional) in the backend and executed by citizen users via the app itself. A wide variety of task types, e.g., reported problem validation, route tracking and description or data input and surveys, make it easy to obtain qualitative data.

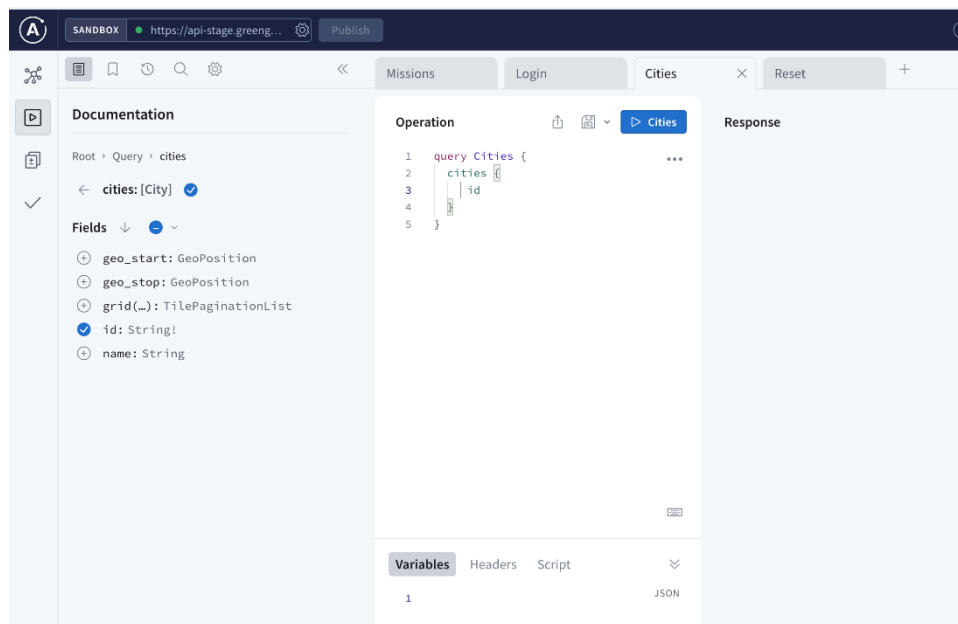


Figure 13: GraphQL view of the GREENGAGE app.

On the other hand, the data can be viewed and analyzed via an open interface. An interactive tool helps to create corresponding queries and run independent evaluations. The API interface can currently be accessed via <https://api-stage.greengage.dev/> <https://api-v2.greengage.dev/>. The admin console is reachable at <https://console-stage.greengage.dev/>. Individual system components (e.g., login and user system) can be maintained and updated adaptively since the microservice architecture concept is used.

Clear distinction needs to be made between the user groups for the app. First, the app is capable of ingesting defined research design in the form of missions for the Citizen Observatories (specifically, engaged citizen report groups) to participate in.

This happens exclusively on the Pilot/coordinator side of the project, where the best approach is elaborated, as such the missions are designed by Pilot owners as well as policy makers. With special authorizations, new missions can also be added programmatically and thus tasks can be assigned to users.

Further, the app can be revised into a self-reporting tool for citizens to collaborate with the policy makers by communicating relevant issues and problems within the app in the form of individual spots or elaborating the missions themselves to provide further data and suggestions.

Citizen users can browse the available missions, participate in them using exclusively the app or other technological solutions (if the research design defines so and provides integration with the app), as such, collecting the data necessary for evaluating and supporting hypotheses.

On the other hand, (urban) developers, architects and scientists can use the data to carry out evaluations. The outcoming data comes in a raw format and is not manipulated in any manner, thus needing specifications where to be sent to, so that other tools can derive conclusions from it and visualize it in a comprehensive manner. Optionally, the app can also be customized to ingest visualized data at a later stage to be displayed to the citizen users but this will be a feature beyond the projects lifetime.

By dividing the GEO data into "tiles", spatial delimitations can be carried out. The use of known data schemas (e.g., GeoJSON) ensures interoperability.

Interoperability with other tools:

As the GREENGAGE app positions itself as a citizen-oriented application and links a wide variety of technologies and partners, a high level of interoperability is guaranteed. By using the subgraph technology standards of GraphQL, other data sources (e.g., MindEarth) might also be connected in the future (i.e., after the lifetime of the project). In addition, the use of webhooks during development also emphasizes interlinked data communication.

a) Data Sharing and Integration

Access to data (unless directly user-related) is always guaranteed via the publicly accessible interface. A restriction is only provided where there are security-relevant processes (e.g., passwords etc.).

b) User Experience Continuity:

Instructions for providing a RESTful service in the GREENGAGE app ecosystem are provided in a separate repository.

3.2.3 MODE

MODE software technology uses smartphones to automatically track the distances travelled and transport modes used, thus enabling software developers, system integrators and transport operators to design innovative mobility services. MODE provides the missing link for the development of innovative mobility services: automatic, reliable recognition of the means of transport used, and travelled routes using smartphones.

Role within the Citizen Observatory:

MODE is part of the GREENGAGE app and is used transparently to record user trips (including transportation mode) for selected trips. Due to battery and privacy considerations, it is not feasible to track every single trip of every user, but CO shall decide which trips are deemed worthwhile to be tracked as users will have to actively start tracking of those trips.

MODE is part of the data collection area. Acquired data can be analyzed to assert if e.g., a given policy change influences the modes of transport used.

User Interaction and Capabilities: Users will need to actively start and end any tracked trips as part of the GREENGAGE app, so they are aware, that they are tracked for the given trip.

Interoperability with other tools:

Interoperability is mostly meant to be used with the GREENGAGE app, but any other tool could access the MODE service if needed.

a) **Data Sharing and Integration:** Interoperability will be via a dockerized REST service which gets raw trip input from smartphone apps, and outputs the trips including transportation mode in (Geo)JSON format. These trips can then be visualized in specific software like kepler.gl¹⁸ or UrbanTEP/VISAT.

b) **User Experience Continuity:** The tool (library) itself does not have a graphical user interface and hence is not visible to the user, except for the Tracking buttons. It is integrated into the GREENGAGE app and is used seamlessly.

¹⁸ <https://kepler.gl/>

3.2.4 IoT Sensors

HOPU provides tools that consists of IoT sensors to collect different kinds of data, for example air quality data, noise pollution and presence/crowd monitoring. Each one of these types of data corresponds with a different sensor with its characteristics:

Air quality (Smart Spot): Smart Spot are configurable IoT device that allow monitoring of different environmental factors, such as air quality (gases and suspended particles), temperature, humidity and noise, as well as integrating weather stations. The inclusion of all these sensors and capacities in a single device provides savings in the installation, maintenance and management as well as in communications. As for the connection possibilities of this device, it is offered in multiple versions, including Wi-Fi, LoRa, GSM/GPRS and NB-IoT.



Figure 14: Smart Spot device

Libelium One Sound Level: Libelium One IoT sensor is another wireless IoT gateway designed for continuous monitoring of a huge range of parameters covering the most relevant IoT applications, for example noise pollution. Thanks to automatic sensor detection, no programming is needed for deployment. Easy and quick installation on walls or poles in combination with a solar panel to maximize its performance. The Noise Level sensor for One has an integration window size from 2 - 15 mins with 1Hz of sample rate and 1 minute of window size with 8Hz of sampling rate. Each measurement has inherent sound pressure Level with the A-weighting filter in dB.



Figure 15: Libelium One Sound Level device

However, due to budgetary constraints and the specific needs of the pilots, low-cost mobile sensors were required. Consequently, HOPU focused on identifying off-the-shelf, affordable, and easily deployable sensors, such as the Atmotube PRO, which proved to be valuable for the project.

Atmotube PRO is a small and compact wearable sensor which can collect data on PM1, PM2.5, PM10, VOCs (Volatile Organic Compounds) and also weather conditions like temperature, humidity and pressure. This sensor is designed to provide the information related the air quality (by air quality index in the range 0-100) to the citizen in real time and on the move (e.g., while walking).



Figure 16: Atmotube PRO wearable sensor

Interoperability with other tools:

Regarding the interoperability with other GREENGAGE tools, all these sensors could be integrated with the GREENGAGE app, being able to visualize the data reported by the sensors over time or at each of the locations, giving the citizen an insight into the environment near him.

Smart Spot and Libelium One device (HOPU sensors) could connect by MQTT broker to the GREENGAGE architecture and data post-processing. In general terms, this data is sent in JSON format, whereas Atmotube wearable sensors send data collected by the device to a cloud storage and the access to data is facilitated by an Cloud API and HTTPS request that returns a JSON or a CSV file with the pollutant's measurements, timestamp and the location where measure has been taken. The process to integrate this kind of data on GREENGAGE architecture and other GREENGAGE tools is the same as Smart Spot and Libelium One. Integration with the Atmotube API was performed during second iteration in order to distil air quality data gathered during GREENGAGE campaigns.

3.3 Area of Concern 3: Analysis and Visualization for Insights Generation

Area of Concern 3 of the GREENGAGE Citizen Observatory Community Journey, titled "*Analysis and Visualization for Insights Generation*", represents a crucial stage in transforming the data collected from earlier areas into actionable insights. This area of concern is where the raw data, gathered through the efforts of citizen observers and various data collection tools, undergoes a transformation into meaningful information. It involves the use of sophisticated analysis and visualization tools to interpret and present the data in a form that is comprehensible and useful for decision-making.

In this area of concern, users are expected to engage with various analytical tools and platforms. The process begins with data integration, where tools like Apache NiFi play a role in extracting and processing the collected data. Following this, data storage solutions like Apache Druid ensure secure and efficient data management. The critical aspect of this area is the visualization and interpretation of data, facilitated by tools like Apache Superset and UrbanTEP/VISAT. These tools help in creating advanced visual representations of the data, making it easier to understand and communicate findings to a broader audience.

Furthermore, this area of concern is instrumental in democratizing data access, as tools like Datahub offer a centralized point for all users to access and analyze the data. An innovative aspect is the use of

tools like DigitalTwin, which create dynamic virtual models of urban environments, integrating data from previous areas of concern and enabling the simulation and analysis of various urban policies and actions.

In essence, Area of Concern 3 is about turning data into decision-making ready information. It is where the efforts of citizen observers culminate in insights that can guide policy decisions and foster a deeper understanding of environmental and urban challenges. The expectation is that users, equipped with the insights generated in this area, can contribute more effectively to sustainable urban development and environmental stewardship.

3.3.1 Apache NiFi

Apache NiFi¹⁹ is an open-source data integration tool designed to automate the flow of data between systems in a secure, controlled, and extensible manner (Figure 17). It serves as a powerful platform for orchestrating data movement and supporting a wide range of data integration scenarios. NiFi operates through a web-based user interface that allows users to design and manage data flows through a visual representation of the process.

NiFi excels in ingesting, transforming and transferring data across various sources and destinations. Regarding to the **data ingestion**, NiFi facilitates the ingestion of data from different sources, such as databases, IoT devices, log files, and more. It supports a variety of protocols, including HTTP, MQTT, and JDBC, enabling seamless connectivity with diverse systems.

The main component of NiFi is the **processor**. Processors allow users to **enrich, filter, and manipulate data** on the fly. Among the available processors, those for data acquisition (e.g., ConsumeElasticSearch, QueryDatabaseTable, ConsumeKafka, GetHDFS, GetMongo, FetchFile, FetchS3Object, GetHTTP), for data transformation (e.g., ConvertRecord, ExtractText, ReplaceText, FlattenJson, SplitJson, JoltTransformRecord, MergeContent, ExecuteScript) and for data storage (PutDatabaseRecord, PublishKafka, PublishMQTT, PutEmail, PutFile, PutMongo) can be found. In case a custom processor is needed, NiFi provides the capabilities for its development.

¹⁹ <https://nifi.apache.org>

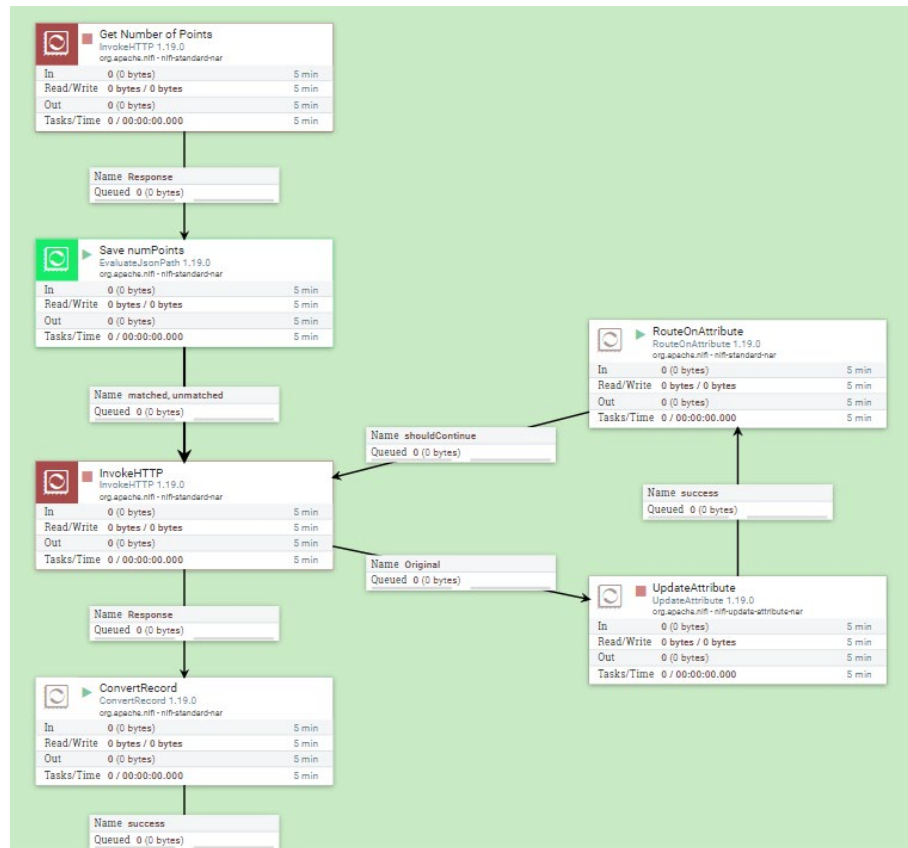


Figure 17: Example of a NiFi dataflow.

NiFi allows routing the data through the designed flow according to different characteristics of the data thanks to different processors (e.g., RouteOnAttribute, RouteOnContent). This feature allows developing complex data flows to fit user needs.

In addition to the wide variety of processors, one of the main strengths of NiFi is its web-based interface. Thanks to this web interface, users do not need to spend time learning complex flow definition languages. The design of the flow is intuitive through the drag-and-drop of the different processors and its connections on the work grid.

In addition to the data itself (payload), each FlowFile at NiFi contains a set of key-value pair representing its metadata. This metadata represents all attributes of the data, e.g., the creation date of the data. In addition to this metadata, NiFi includes detailed data provenance tracking, allowing users to trace the origin, transformation, and destination of each piece of data. For that, NiFi provides the Provenance Repository component, in which the history of each FlowFile is stored. This history is used to provide the Data Lineage of each piece of data.

Regarding to scalability of the system, NiFi is designed to scale-out through its deployment in a cluster of computers. NiFi provides the mechanisms to transparently synchronize FlowFile queues among different nodes on the cluster and to properly balance the load among them.

Last, NiFi provides the proper mechanisms to ensure the security and governance of the data. NiFi incorporates robust security features, including encryption, authentication, and authorization mechanisms. It ensures the multitenancy and collaborative development of data flows.

In summary, Apache NiFi serves as a versatile and robust data integration tool, empowering the GREENGAGE project with efficient data movement, transformation, and management capabilities. Its unique features and scalability make it an asset in orchestrating the complex data workflows required for the success of the Citizen Observatory.

Role within the Citizen Observatory

Apache NiFi plays a pivotal role within the Citizen Observatory, contributing significantly to the “Analysis and Visualization for Insights Generation” area of concern. Thanks to its wide variety of processors for ingesting data, NiFi facilitates the seamless integration of data from diverse community sources into the observation process. It ensures that data generated through community engagement activities, such as app generated by mobile apps, surveys or collaborative platforms, is efficiently ingested and prepared for further processing.

NiFi acts as a central hub for collecting, filtering, and curating data contributed by citizens. In addition to its ability to ingest data from various source, it can apply transformations, and route information based on predefined conditions, ensuring that crowdsourced data is curated in real-time. NiFi can become a fundamental tool for maintaining data quality, relevance, and consistency.

In a nutshell, NiFi serves as the backbone for data movement and transformation, enabling the efficient transfer of curated data to tools focused on analysis and visualization. By automating the flow of data between different analytics and visualization platforms, NiFi ensures that insights derived from one tool can be seamlessly integrated into the analytical workflows of subsequent areas. This contributes to a cohesive and integrated approach to deriving meaningful insights from the observed data.

On the other hand, NiFi provides a user-friendly web-based interface that empowers users to design, monitor, and manage data flows visually. Users can interact with the tool to create, modify, and optimize data pipelines without the need for extensive coding or scripting. This visual representation enhances the accessibility of the tool across diverse team members, including those with varying technical expertise.

Interoperability with other tools

Apache NiFi excels in interoperability, seamlessly integrating with other tools within the Citizen Observatory to facilitate cohesive data sharing and processing. As mentioned before, NiFi enables the ingestion of data from various sources and can seamlessly share this data with other tools in the project. For example, data collected from sensors, mobile apps or community platforms can be efficiently ingested by NiFi and shared with analytical tools for further processing. NiFi supports a wide array of integration protocols, including REST, HTTP, MQTT, and more. This flexibility ensures compatibility with diverse tools and systems within the project. NiFi's data transformation capabilities are crucial for ensuring that data is in a compatible format for downstream tools. Whether it is preparing data for visualization or analysis, NiFi can apply transformations to align the data with the requirements of specific tools, promoting seamless integration.

Regarding the user experience continuity, NiFi provides a visual interface for designing data flows, enhancing the user experience. The visual representation of data pipelines remains familiar, regardless of whether the user is working on data ingestion, transformation, or routing. This ensures that users can easily navigate and manage data flows throughout the project. The provenance tracking aligns with the project's need for transparency and traceability. The metadata associated with each data flow, including its origin, transformations applied, and destination, contributes to a comprehensive understanding of the data's journey. This integrated approach supports a seamless user experience by providing context and insight into the data's history.

3.3.2 Apache Druid

Apache Druid ²⁰ is an open-source data store designed for high-performance real-time analytics. It excels in ingesting and querying large volumes of data with low latency, making it ideal for interactive applications. Druid is commonly used for business intelligence, operational analytics, and handling time-series data. Its architecture allows for scalable, fast, and efficient data handling and is particularly effective in environments where real-time insights and quick query responses are crucial. Druid is well-suited for use cases that require rapid aggregation and flexible data exploration across large, dynamic

²⁰ <https://druid.apache.org>

datasets. Figure 18 shows the data source view in which users can check the columns and datatypes a data source holds.

_time	Site_ID	NOx	NO2	NO	PM10	O3	Temperature	Objectid	Objectid2	NPM10	VPM10
1993-01-04T00:00:00.000Z	188	27	27	0	31	48	empty	empty	815,518	empty	empty
1993-01-04T01:00:00.000Z	188	38	36	1	30	42	empty	empty	1,438,309	empty	empty
1993-01-04T02:00:00.000Z	188	32	31	1	20	44	empty	empty	1,159,358	empty	empty
1993-01-04T03:00:00.000Z	188	32	31	1	31	42	empty	empty	1,188,550	empty	empty
1993-01-04T04:00:00.000Z	188	17	17	0	26	48	empty	empty	1,188,548	empty	empty
1993-01-04T05:00:00.000Z	188	27	25	1	29	42	empty	empty	826,086	empty	empty
1993-01-04T06:00:00.000Z	188	36	31	4	33	32	empty	empty	815,514	empty	empty
1993-01-04T07:00:00.000Z	188	124	63	40	31	14	empty	empty	86,779	empty	empty
1993-01-04T08:00:00.000Z	188	325	73	165	39	10	empty	empty	439,636	empty	empty
1993-01-04T09:00:00.000Z	188	686	111	376	79	12	empty	empty	1,188,546	empty	empty
1993-01-04T10:00:00.000Z	188	947	164	513	111	14	empty	empty	815,510	empty	empty
1993-01-04T11:00:00.000Z	188	659	149	334	88	12	empty	empty	1,188,544	empty	empty
1993-01-04T12:00:00.000Z	188	229	82	96	52	18	empty	empty	815,506	empty	empty
1993-01-04T13:00:00.000Z	188	227	76	99	30	22	empty	empty	86,778	empty	empty
1993-01-04T14:00:00.000Z	188	222	65	103	35	18	empty	empty	439,635	empty	empty

Figure 18: Apache Druid data source view.

Apache Druid offers a broad set of features that fulfil several requirements of the GREENGAGE project regarding data management:

- **Real-Time Data Ingestion and Querying:** Druid's prowess in handling large volumes of data in real-time is a vital asset for the GREENGAGE project. In an era where immediate insights from environmental data are critical, Druid's ability to ingest and query data in real-time ensures that stakeholders can access up-to-the-minute information. This feature enables timely responses to changing environmental conditions, which is essential for informed decision-making.
- **Scalability:** In the long term, the GREENGAGE project will generate and processes vast amounts of data, and Druid's scalable architecture is well-suited to meet these demands. As the project evolves and data volumes grow, Druid can seamlessly expand its capacity to ensure consistent performance. This scalability is crucial for accommodating the evolving data needs of the project and preventing bottlenecks in data processing.
- **Time-Series Data Analysis:** Environmental monitoring inherently involves tracking and understanding trends over time. Druid's strength in time-series data analysis aligns perfectly with the project's objectives. It empowers users to delve deep into historical data, identify patterns, and gain insights into how environmental factors evolve over various timeframes. This capability is essential for making data-driven decisions and understanding the long-term impacts of environmental changes.
- **Flexible Data Aggregation:** The GREENGAGE project will often require rapid data aggregation for quick analysis and reporting. Druid's ability to perform rapid data aggregation supports this need. Whether summarising data from various sources or generating real-time aggregations, Druid offers the flexibility required to extract meaningful insights efficiently. This feature streamlines the data analysis process, allowing stakeholders to focus on interpreting the results and deriving actionable insights.
- **Interactive Dashboards:** Visualizing complex environmental data in an accessible format is crucial for effective communication and decision-making. It empowers users to create dynamic and interactive visualizations that facilitate data exploration. Although this is not Druid's primary role in the project, it enables users to do quick analysis or preview whether data has loaded correctly (Figure 19).

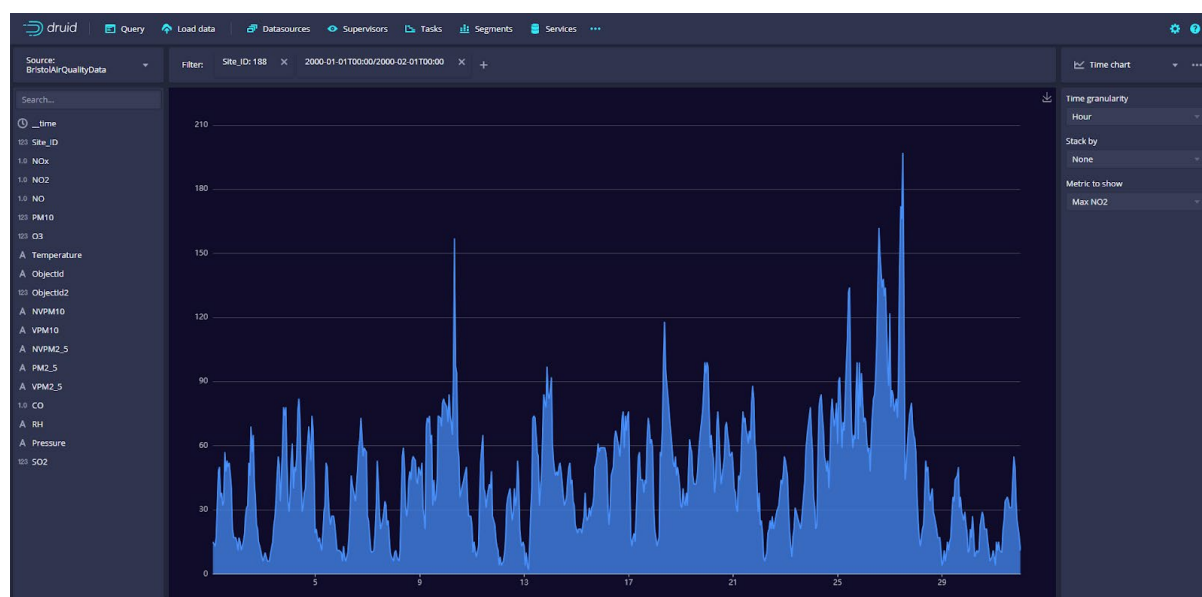


Figure 19: Apache Druid analytics view.

Apache Druid's robust feature set meets the requirements of the GREENGAGE project regarding data management and aggregation. Its real-time data capabilities, scalability, time-series data analysis, flexible data aggregation, and support for interactive dashboards make it an invaluable tool for managing, analysing, and gaining insights from environmental data. These features empower project stakeholders to host and manage data, which can then be connected to other tools to be exploited appropriately.

Role within the Citizen Observatory

Within the GREENGAGE project, Apache Druid assumes a pivotal role in the '*Analysis and Visualization for Insights Generation*' area of concern, whose capabilities are instrumental in harnessing the full potential of the project's large-scale datasets. Druid's primary contribution revolves around its prowess in efficiently managing extensive volumes of data and facilitating rapid data access, which are indispensable requirements for this critical area.

Apache Druid empowers users to interact with the tool, offering a comprehensive set of data processing capabilities that streamline the analysis of time-series data—an essential component for gaining insights into evolving environmental trends. Its ability to handle time-series data efficiently allows stakeholders to track and understand changes over various temporal scales, from short-term fluctuations to long-term patterns. This capacity for in-depth time-series analysis is invaluable for the GREENGAGE project, where the ability to decipher environmental trends is central to achieving its objectives.

Interoperability with other tools

Apache Druid seamlessly integrates with other tools within the GREENGAGE Citizen Observatory to facilitate data sharing and integration. For instance, data collected from sensor tools like IoT of-the-shelf sensors, MODE, MindEarth or GREENGAGE app can be efficiently imported into Druid for efficient storage. Furthermore, by creating NiFi flows or scheduled tasks, data from Open Data portals may be imported into it. Data can flow smoothly from data collection tools to Druid, ensuring that insights from various sources can be easily combined and explored. The only requirement is that the data hosted in Druid must be in a columnar format or should be transformable to it.

Apache Druid contributes significantly to a seamless user experience across different areas of concern of the GREENGAGE project. Insights gathered from initial data collection tools, such as sensor data and mobility patterns from MODE, can be further stored and visualized within Druid. However, given its role as the primary data repository, maintaining the integrity and accuracy of the tool is of utmost importance. To minimise the risk of errors and data discrepancies, access to Druid will be limited to individuals who possess the requisite skills and knowledge in data processing and management. This selective approach

ensures that the data hosted within Druid remains reliable and consistent throughout the project's lifecycle, aligning with GREENGAGE's commitment to high data quality standards.

3.3.3 Apache Superset

Apache Superset²¹ (or Superset) is an open-source data visualization and data exploration platform designed for modern data teams. It empowers users to create and share interactive dashboards and engaging visualizations, offering a rich set of features to explore and understand large and complex datasets. Developed under the auspices of the Apache Software Foundation, Superset integrates effortlessly with most SQL-speaking databases, making it an incredibly versatile tool in the world of big data analytics.

Superset is particularly well-suited for creating charts, dashboards, and conducting SQL queries to datasets. Its user-friendly interface enables even those without extensive coding experience to build informative visualizations (Figure 20). This accessibility is combined with powerful analytical capabilities, allowing users to create custom visualizations, apply filters, and drill down into specific data points. The tool supports a wide range of chart types, from basic line and bar charts to more complex types like geospatial charts, making it a flexible choice for various data visualization needs.



Figure 20: Example of Apache Superset Dashboard.

One of the primary features of Apache Superset is its intuitive and responsive dashboard creation interface. Users can drag and drop components to build dashboards that present real-time data insights. Additionally, Superset provides robust SQL IDE support, enabling data professionals to write, run, and visualize the results of SQL queries directly within the platform as shown in Figure 21. This feature is particularly useful for conducting deep data analysis and troubleshooting.

²¹ <https://superset.apache.org>

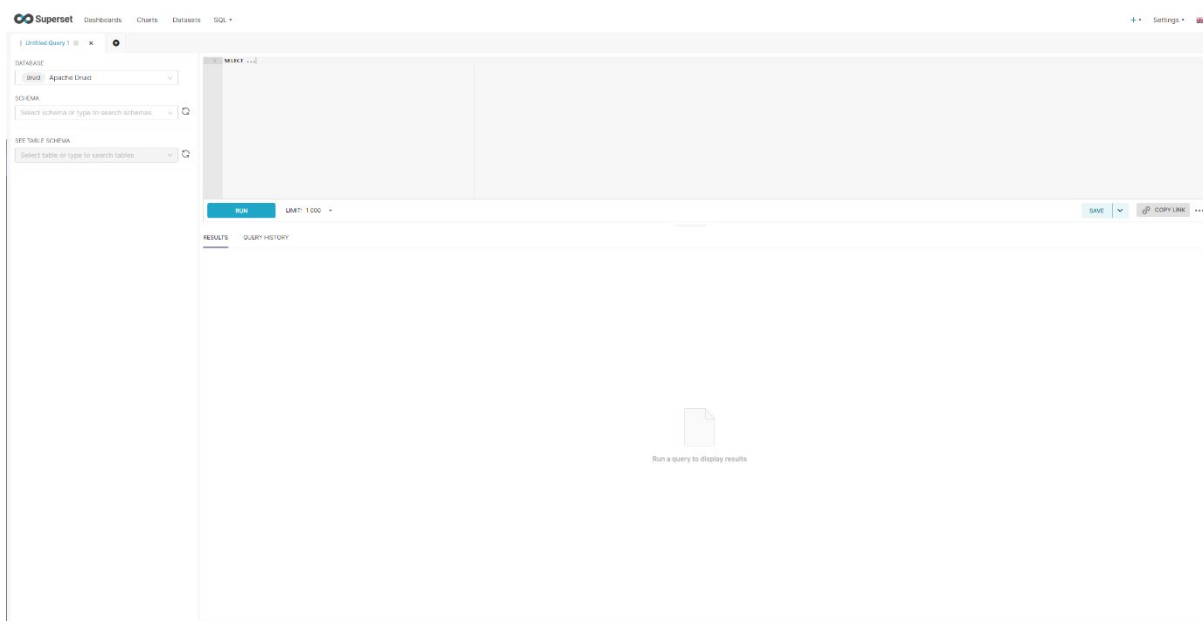


Figure 21: SQL lab from Apache Superset.

Unique to Apache Superset is its extensibility and customization. It offers a wide array of customization options, from changing themes to adding custom visualization plugins, and allowing teams to tailor the tool to their specific needs. On the technical side, Superset is built to handle large datasets efficiently. It leverages caching mechanisms to speed up query performance and can integrate with most SQL database backends, including popular choices like PostgreSQL, MySQL, and SQLite, as well as more specialized data stores like Apache Druid and Google BigQuery.

Role within the Citizen Observatory

In the GREENGAGE project, Apache Superset's role is pivotal in the 'Analysis and Visualization for Insights Generation' area of concern. This tool enhances the project by facilitating sophisticated data analysis and the creation of dynamic, interactive visualizations. Through its user-friendly dashboards and exploration tools, Superset transforms intricate datasets into clear, actionable insights, essential for informed decision-making in environmental monitoring. Its capacity to generate detailed reports and tailor data presentations makes it an asset for community engagement, underpinning the project's goal of effective environmental management and policy development.

Within the GREENGAGE project, Apache Superset demonstrates remarkable capabilities in dashboard creation, combining charts from varied data sources and incorporating user-conducted analyses. Its interactive controls allow stakeholders to present insights derived from heterogeneous data effectively. These dashboards are not just visually engaging but also highly informative, facilitating a deeper understanding of environmental trends and patterns among diverse project participants.

To streamline the process and enhance user experience, the connection of Superset with various data sources is managed by Pilot support teams. This approach alleviates the complexity for end-users, ensuring that they can focus on data analysis and interpretation without being burdened by the intricacies of data integration. This thoughtful arrangement allows stakeholders to engage with the data more efficiently and effectively, enhancing the overall impact of their contributions to the GREENGAGE project.

Interoperability with other tools

Apache Superset's interoperability within the GREENGAGE Citizen Observatory is critical as Superset must ingest data from the other tools available in the GREEN Engine. Apache Superset can ingest datasets from over 40 sources, including the most common relational databases as well as some specialized ones.

It can effectively share and receive data from various tools used in the *Data Crowdsourcing and Curation* area of concern, such as sensors or data collection applications through the Druid tool. This capability allows for the visualization and analysis of data collected in earlier project areas. Furthermore, Superset enhances the user experience by allowing insights gathered from one phase or tool to be further explored and visualized, providing a seamless transition and deeper understanding as users move through different project stages.

3.3.4 Data Quality Dashboard

In Area of Concern 3, the Data Quality Dashboard (DQD) plays a supportive role within the Citizen Observatory, contributing significantly to the overarching objectives of this area. Tasked with supporting the initial quality assessment of acquired data, the DQD scrutinizes the quality of collected data and resolves detected issues, validating data correctness and enhancing trust levels for informed decision-making. The DQD is recommended as an entry point for subsequent investigations within the project when applicable.

The DQD is tailored to address various data types and sources, including time-series-based air quality, noise, and meteorological data, as shown in Figure 22. Given the inherent quality challenges associated with time-series data, particularly sensor data, the DQD becomes instrumental in identifying issues at different stages of the measurement gathering workflow. These challenges encompass problems related to power supply, communication persistence, sensing circuit flaws, data recording and storage, and faulty installation of sensing devices. Specifically, the DQD facilitates the identification of data quality issues by enabling users to assess the completeness of time-series data, detecting missing values and periods. It also ensures the accuracy of time-series data by identifying duplicate timestamps and conducts anomaly detection, flagging outliers or unusual sequences of identical values. The tool adheres to industry standards and provides users with overall assessed data quality scores.

The Data Quality Dashboard is built on a modern React/TypeScript stack with Plotly.js for advanced visualizations. It connects to Apache Druid as its backend data store, enabling efficient handling of time-series data. The dashboard implements four dimensions of data quality: completeness (tracking data gaps and missing values), validity (enforcing value ranges and constraints), uniqueness (measuring duplicate values), and accuracy (via statistical analysis and trend detection). The tool provides both automated scoring mechanisms and interactive visualizations, including calendar heatmaps for temporal analysis (with the focus on pattern analysis and data gaps detection), trend lines (also for pattern and data gaps detection), histograms for distribution analysis (and outlier detection by linked view with map and trend line plots), and map plots for geographical data visualization. It supports various data aggregation intervals (15min to daily) and implements data throttling to handle large datasets efficiently.

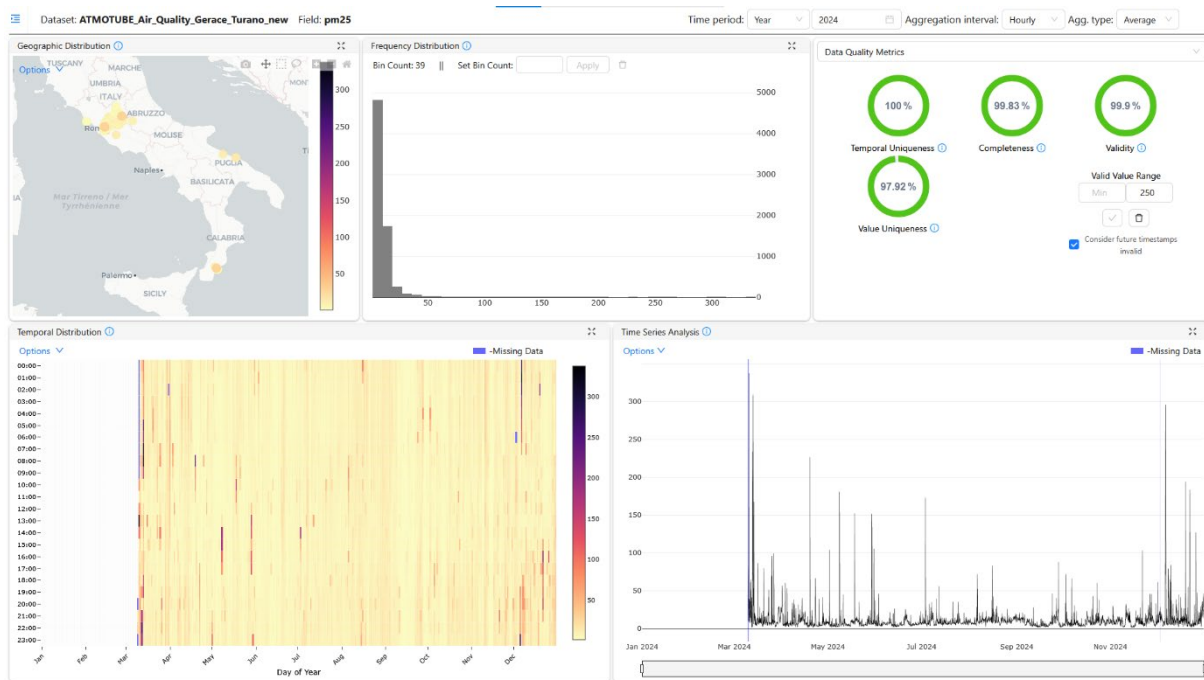


Figure 22: Data Quality Dashboard.

Users engage with the tool by directly interacting with the diverse elements featured on the dashboard, including graphs, selection boxes, and fields. This hands-on approach empowers them to refine visualizations and conduct analyses focused on specific variables or desired timeframes. The essence of the dashboard's graphical user interface (GUI) lies in the interactive nature and responsive behavior exhibited by its elements.

Interoperability with other tools

The DQD is poised to seamlessly integrate with GREEN Engine's data infrastructure. Through harmonization with the central data repository (Apache Druid) users will have the capability to effortlessly access and work with all pertinent datasets stored within the repository. This interoperability ensures a streamlined process for visualizing and analyzing timeseries data collected from sensor tools or any other relevant data sources related to Citizen Observatories.

3.3.5 UrbanTEP / VISAT

In the context of the GREENGAGE project, GISAT provides an open-source, web-based application built using its Visual and Analytical Toolboxes (VISAT) — a continuously evolving visualization framework previously applied in various implementations for clients such as the World Bank Group (UrbanTEP / PUMA), ESA (UrbanTEP, WorldCereal, WorldWater), and NordRegio (NordMap) (see Figure 23).

During the second iteration, UrbanTEP / VISAT has placed a strong emphasis on storytelling with geospatial data. The platform enables users to access interactive narratives that combine maps, charts, and descriptive text, making complex data accessible and engaging for a wide range of audiences. This storytelling approach supports knowledge sharing, decision-making, and public engagement by allowing users to view and share data-driven stories about urban environments and sustainability challenges.

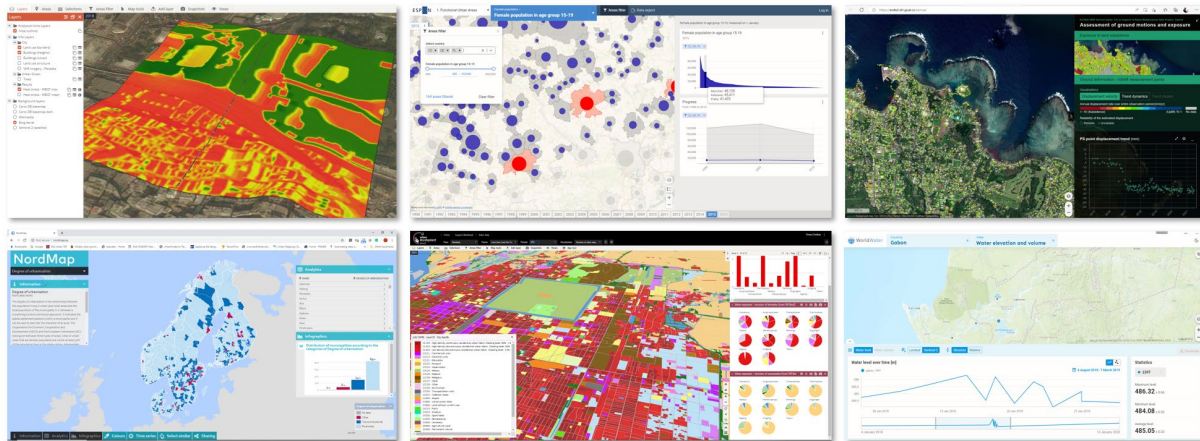


Figure 23: Examples of GISAT's web-based analytical applications with VISAT framework application.

The **Urban Thematic Exploitation Platform (UrbanTEP)** leverages modern information technologies to bridge the gap between vast Earth Observation (EO) data collections and the growing demand from science, planning, and policy for up-to-date insights into the status, characteristics, and evolution of urban systems.

A key component of the modern **UrbanTEP / VISAT** platform used in the GREENGAGE project is an **open, web-based, and participatory application**. It is built on an updated version of the VISAT framework with newly updated technical architecture, utilizing the latest libraries and visualization methodologies. This enables:

- **High-performance data access, analysis, and visualization**, with improved flexibility and enhanced support for server-side rendering (SSR) features, ensuring faster and more efficient delivery of visual content.
- **User-oriented customization**, allowing tailored development and sharing of data products (collected surveys, Remote Sensing data, GREENGAGE stakeholders' data)
- A visually consistent **design aligned with the GREENGAGE project website**, tailored to communicate effectively with diverse user groups, **promote the project's visibility**, and **encourage participation in Citizen Observatories**.

VISAT's intuitive interface allows users to assemble visualizations, text, and multimedia into coherent storylines. Features such as story maps, guided data explorations, and interactive dashboards empower both experts and non-experts to communicate findings, illustrate trends, and support participatory decision-making. The application's design encourages collaboration and sharing, making it a valuable tool for Citizen Observatories and stakeholder engagement. The overarching objective of UrbanTEP / VISAT is to enable non-expert users to easily generate actionable indicators and information for effective sustainable urban development support based on a joint analysis of various data sources such as official survey data, EO mission data, socio-economic statistics, and data collected via social media or Citizen Science. High volume of data can be processed, collected and analyzed by UrbanTEP / VISAT to provide a broad spectrum of urban information products and related services for visualization and analytics that have yet successfully been used by more than 240 institutions (science, planning, NGOs, policy) from 41 countries (i.e. World Bank Group, United Nations, Organisation for Economic Cooperation and Development, World Food Programme, Bill and Melinda Gates Foundation, Group on Earth Observation, Global Platform for Sustainable Cities).

Urban Thematic Exploitation Platform (UrbanTEP) is currently a system environment composed of several sub-components:

- A web application, interacting with a rich EO data and derived datasets catalogue associated to a data storage, as well as external WPS services developed by various providers,
- An updated Visualization and Analytics Toolbox, that serves for data exploitation (VISAT),
 - Access to VISAT's updated backend that handles data loading, allowing to query and communicated with metadata service and deliver data directly to client by handling data transformations and potential error handling.

- Access to newly developed VISAT's metadata service that handles data operations, preparation and caching.
- Newly available VISAT's frontend library that allows to handle delivered visualizations and ensure the result will be in a user-friendly interpretation.

For example, policymakers can use UrbanTEP / VISAT to access a story map that visualizes changes in green spaces and air quality over time, combining satellite imagery, survey data, and citizen observations. This story can be shared with the public or decision-makers to foster dialogue and inform policy. Citizens can use UrbanTEP / VISAT to communicate directly with local government, providing evidence and feedback based on geospatial data and personal observations. Additionally, experts can leverage VISAT as a reference platform, similar to tools like Apache Superset, to access data directly and perform advanced analyses according to their own methodologies.

The UrbanTEP's visualization, exploration, and analytical module (VISAT), developed by GISAT, is a powerful web-based application tool designed to facilitate data analysis and visualization of geospatial data. The solution is scalable and supports multiple standard data formats, including GeoJSON, Shapefiles, WMS, GeoTIFFs, and more. The framework is also capable of easily implementing new data types and working with newly proposed metadata structures.

UrbanTEP also offers robust data curation and storage capabilities using various database technologies such as SQL and NoSQL, as well as Open Data like Copernicus datasets. The platform can work with different types of data, including raster, vector, and tabular data, and can perform geospatial analyses.

Visualization and Analytics Toolbox (VISAT) is based on an open extendable constantly developing framework, designed as user-friendly and intuitive interface and allows data to be presented using dynamic and interlinked map and non-map component (e.g., charts, tables and graphs). The platform features two interfaces, a Web GUI, and a REST API that handles communication between multiple parts of the application, including frontend and backend.

Backend services - set of microservices, written in TypeScript for the NodeJS environment and dockerized for easier maintenance and better scalability. At the same time, storage technologies are used according to the nature of the data + own spatial data in PostgreSQL / PostGIS. Complex metadata structures and dynamic database schemas are handled using MongoDB.

Frontend components - a set of packages written in React, using Redux for state management, D3JS and Nivo Charts for graph visualization and DeckGL for map elements.

The deployment strategy of the UrbanTEP is characterized as a Cloud solution (SaaS) and can run on various platforms that support the Docker technology. The GISAT's framework, used for VISAT creation, is based on open-source development dependencies and includes hundreds of packages that are constantly updated by third-party developers. The most common technologies and dependencies used are NodeJS, Postgres, MongoDB, and S3. Overall, the VISAT is a reliable and efficient tool for geospatial data analysis and visualization. The overall technical architecture of the tool is displayed in Figure 24.

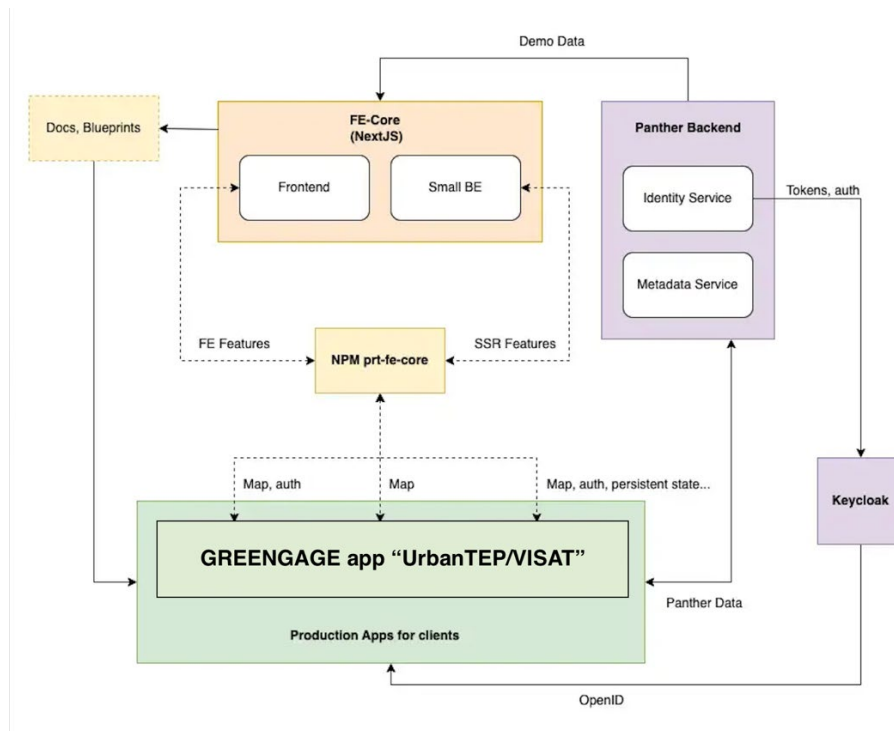


Figure 24: Overall technical architecture of the web application.

VISAT has been recently updated including its backend and frontend services. GREENGAGE development will already leverage these improvements (see Figure 25).

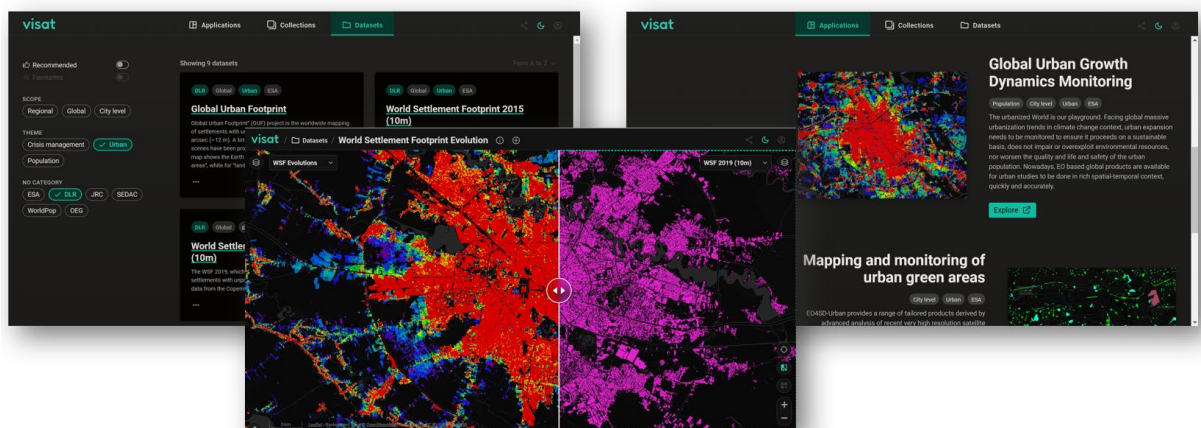


Figure 25: Example of upgraded UrbanTEP's VISAT GUI interfaces.

3.3.6 DataHub

DataHub²², in the context of the GREENGAGE project, is an advanced data management platform that plays a crucial role in integrating and organizing diverse data streams. Its primary function is to aggregate data from various sources, providing a unified, accessible environment for data analysis. DataHub excels in metadata management, a critical feature that adds context and understanding to the data, crucial for informed decision-making. The platform also offers powerful data discovery and exploration tools housed within a user-friendly interface (Figure 26), making it accessible to users of varying technical expertise. DataHub is technically scalable, ensuring consistent performance as data volumes grow and adhering to strict data security and compliance standards. In GREENGAGE, DataHub's ability to provide a centralized data repository is vital for supporting advanced data analysis and enhancing the decision-making process in environmental monitoring and policy development.

²² <https://datahubproject.io>

Throughout various stages of data-centric projects, numerous challenges and queries arise. DataHub is designed to seamlessly guide participants through every phase of these projects, from the initial planning and data sourcing to the application of advanced data analytics and insights extraction. It offers comprehensive guidance for data management and utilization and practical advice, templates, and tools tailored to enhance collaborative efforts.

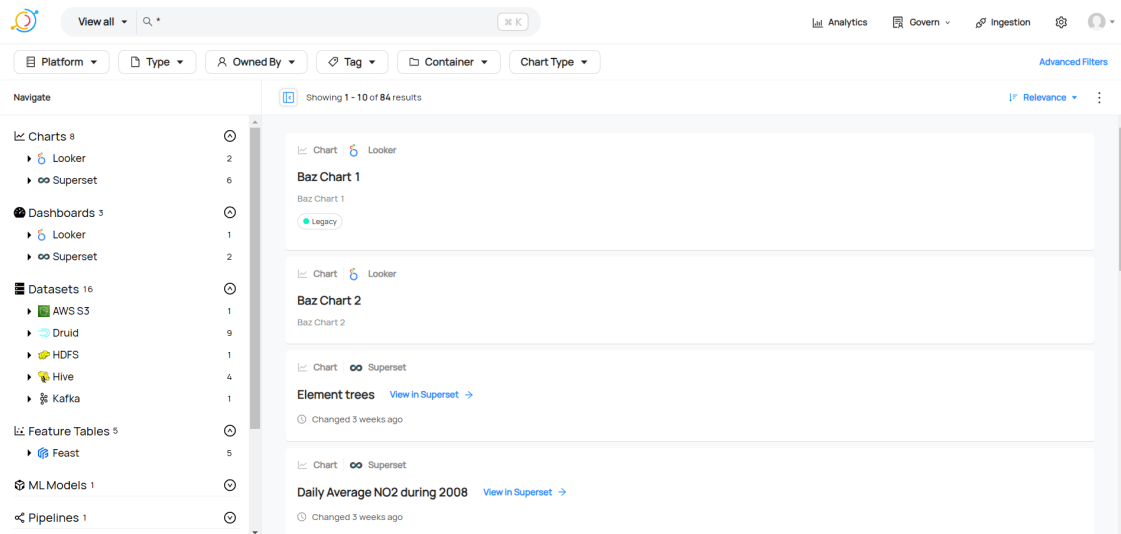


Figure 26: Explore view of DataHub.

Role within the Citizen Observatory

DataHub plays a multifaceted role in the GREENGAGE project, acting as a central data aggregation, management, and analysis hub. Its functions include aggregating diverse data sources for accessibility, managing metadata to add context to raw data, and facilitating data discovery with a user-friendly interface. The scalability and robust security of DataHub's architecture ensure reliable performance and data integrity, which are crucial for handling growing environmental data volumes. In GREENGAGE, DataHub is instrumental in enhancing decision-making processes for environmental monitoring, underpinning the project's data-driven approach.

- **Aggregation and Accessibility:** At its core, DataHub is tasked with the job of aggregating data from over 50 sources, thus offering a unified and accessible environment for data analysis. This functionality is pivotal as it allows datasets from disparate sources to be aggregated in one place, allowing stakeholders to identify them quickly and easily.
- **Metadata Management:** The brilliance of DataHub lies in its metadata management capabilities. By adding context and understanding to raw data, it elevates the information from mere numbers to actionable insights. This aspect is invaluable, especially when making informed decisions that could have far-reaching environmental implications. Furthermore, Datahub allows versioning of the assets and metadata, which is crucial for reproducing the proposed workflows.
- **Data Discovery and User Accessibility:** The platform is not just about handling data; it is about making data comprehensible and usable. With its powerful data discovery tools and user-friendly interface, DataHub caters to users with varied technical expertise, democratising data analysis.
- **Scalability and Security:** DataHub's scalable architecture ensures consistent performance despite increasing data volumes and adheres to the strictest data security and compliance standards. This scalability is crucial as environmental data continually evolves and expands.

In the GREENGAGE initiative, DataHub serves as the central data catalogue within the GREEN Engine, enabling teams to discover, understand, and govern data assets. In this setup, Apache Druid operates as the metric store, housing detailed time-series and aggregated data, while Apache Superset functions as the primary visualization layer, creating dashboards and charts from Druid's datasets. DataHub ingests metadata from both systems—approximately 80% from Superset and 20% from Druid—cataloguing datasets, dashboards, and charts. This workflow allows DataHub to unify these assets into a searchable catalogue, though current metadata (ownership, tags, glossary terms) is still being enriched to fully support governance. Together, these tools streamline data exploration: Druid supplies analytics-

ready data, Superset visualizes it, and DataHub catalogues and enhances discoverability, ensuring teams can reliably find, interpret, and govern their data resources. However, although its primary role is played in the third area of concern, DataHub is present from creating the communities to explaining the data.

At the outset, in the Community and Co-production Process Management area of concern, DataHub is instrumental. It aids in setting clear objectives for each campaign, identifying necessary datasets, and addressing potential data gaps. Additionally, it facilitates the selection of suitable tools and personnel, ensuring that the data collection process is both practical and efficient.

As campaigns move into the Data collection area of concern, DataHub's strengths become even more apparent. It supports the development of sophisticated data workflows by aiding in tool selection and providing a centralized way of displaying them. Users can visualize how the data is ingested and stored as different versions of one dataset.

Finally, the Analysis and Visualization for Insights Generation area of concern is where DataHub contributes the most with its functionalities. It acts as a showcase of available data with enriched metadata and documentation for each asset. It also provides a centralized location to host charts and dashboards that may be used to present and extract data insights. Overall, DataHub democratizes data access, ensuring that all users have easy access to the information they need. Furthermore, this straightforward information access will allow users to inspire in previous campaigns and develop new ones. This will contribute to the sustainability of the project and the creation of a community of practice.

Interoperability with other tools

Regarding integration and interoperability with other tools proposed in the GREENGAGE project, Datahub enables seamless data sharing and integration. For instance, it can directly import data from IoT portable sensors, from the open data catalogues available in the cities where the Pilots are being conducted or from the databases of MindEarth and GREENGAGE app applications. This contributes to a holistic view of the Pilot and provides a centralized location for analysing the available information. Furthermore, DataHub maintains user experience continuity across project areas. Data from tools employed in the second area of concern can be integrated into DataHub, allowing for deeper insights, correlation and even the creation of charts and dashboards to present it. This seamless integration and continuity ensure that insights from one area are effectively utilized in subsequent ones, reinforcing the project's cohesive and comprehensive Citizen Observatory Community Journey.

In summary, DataHub transcends its role as a mere data management tool. Enriching the GREENGAGE project with its functionalities, it functions as a pivotal hub for data discovery, management, and user empowerment. Its integration across the diverse areas of concern of the Citizen Observatory Community Journey positions DataHub as the central repository, facilitating data localization and enhancing the accessibility and utility of data for all stakeholders involved.

3.3.7 DigitalTwin

DigiTwin, also known as the Argaleo Digital Twin City, is a cloud based digital twin or atlas made by Argaleo²³. This cloud-based platform serves as a data repository and data analysis & visualization tool. It contains different data sets from various sources to visualize the static and dynamic/real-time data for cities, regions, and the whole of the Netherlands. These datasets include different fields such as urban design, mobility, energy, and demography. Users, based on their needs, can request access to a set or all of them and run multi-purpose analyses and 3D visualizations. With the platform itself, users can combine datasets and analyze correlations, which can inform planning decisions.

DigiTwin is a modular web application which can be customized and developed based on the needs and objectives of different users. The main datasets of this application are reliable open-source datasets which are usually provided by governmental organizations such as Netherlands' Cadastre (Land Registry and Mapping Agency), CBS (Central Bureau of Statistics), the Ministry of Infrastructure and Water Management and municipalities. Users, based on their needs and objectives, could link their data sets

²³ <https://www.argaleo.com/en/>

into available and integrated data sets, to perform specific analysis and customize their visualization by adding different layers of information. Users can perform geo-spatial analyses by combining datasets and visualize the results in 3D, 2D or even through tables and charts.

Within the GREENGAGE project, the DigiTwin was used by the North Brabant pilot. It served as data visualization tool and repository of mobility data from the Province of North Brabant. Data gathered with the GREENGAGE app – survey data and GPS photo points on bike path maintenance– during the North Brabant app was integrated into the DigiTwin, to serve as contrast to visualized official provincial datasets. The relation between the DigiTwin and the GREEN Engine is discussed below.

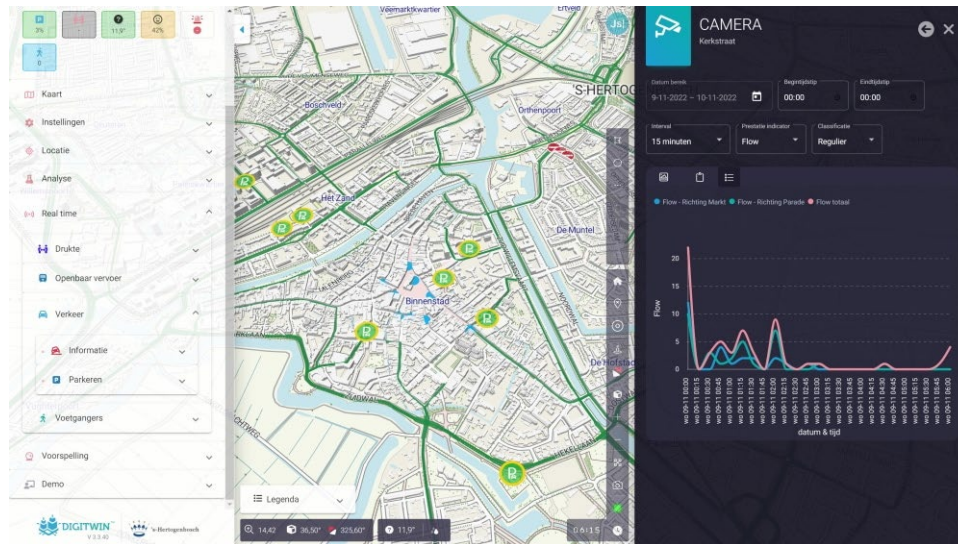


Figure 27: An example of an analysis in the Dashboard

Role within the Citizen Observatory

The DigiTwin is especially used in area of concern 3, when Analysis and Visualization for Insights Generation rear their heads. The platform integrates the gathered data and allows for the combination of the crowdsourced dataset with other spatial planning data. This combination is made visible through isochrones or heatmaps, on a Dutch geospatial map, actively indicating correlations between datasets.

In the GREENGAGE project, the GPS and questionnaires data on bike path maintenance, gathered by the Citizen Observatories are translated into compatible input file formats and integrated into the DIGITWIN datasets. These input files are either made by geolocated images or survey related to specific bike paths, or by using survey and point output from the GREENGAGE app. Next to citizen gathered data from the GREENGAGE app, the DigiTwin is expanded with provincial datasets. Combining the existing embedded datasets of the DigiTwin, such as maintenance planning overviews per path, and physical characteristics, and the user added sets, helps to highlight discrepancies between the subjective appraisal of maintenance and the objective assessment of bike path quality. These observer-identified discrepancies can subsequently be used to optimize the maintenance planning in accordance with the use of the bike path. The DigiTwin allows for the simultaneous overlay of gathered data and historical datasets, for immediate understandable comparisons.

However, citizen observers themselves do not have the possibility to add datasets to the dashboard. There is no direct dataflow between the GREEN Engine apps and the DigiTwin. While the data from the GREENGAGE app can be added for instance, there is no direct link. The collected and cleaned data is submitted by the core group to Argaleo who integrates it. The visualized results are communicated to the citizen observers in interactive sessions in which the sets will be combined with different layers on demand, to explore possible correlations. Furthermore, the dashboard offers the singling out of specific scenarios, initiated by the gathered data, that can subsequently be discussed in person for community engagement.

For citizen observers, the DigiTwin will be made available as a data analysis and visualization tool during the North Brabant datathon. With the citizen-gathered data visualized, and coupled to Provincial datasets on the bike network, the datathon will challenge participants to find fruitful combinations of data to explain discrepancies in maintenance, identify patterns in low appraised areas, and define essential and/or

missing datasets. Outside of the datathon, the DigiTwin is not available for observers. Although they can be involved in the data curation, for many participants from area of concern 2, the visualization developed by experts will be the first mapped result of the gathered data. This then serves as a communication means, as well as visible proof of integration in operational planning contexts.

Interoperability with other tools

The DigiTwin itself is not capable of capturing data, nor does the platform itself curate data. On request, curated datasets can be integrated into the twin. For GREENGAGE this means that the output of citizen gathered data from the GREENGAGE app serves as input for the DigiTwin. There are no other links between GREEN Engine tools planned besides the GREENGAGE app.

The platform can analyze different datasets, with some functions focused on accessibility or range. However, most of the analyses stem from the visual combination of different data sets, which require interpretation. The main role of the DigiTwin then is in Data Exploitation and co-creation. The curation process, as well as the visualized analyses facilitate real life conversation with involved stakeholders – like CO members. Going through the process of exploiting the data for useful insights sets up the possibility for co-creation. This requires taking citizens through the process of data curation, and analysis, to enrich the quantitative visualized results with qualitative additions. As a conversation facilitator, the DigiTwin plays a role in the later steps of the data processing pipeline. Therefore, specifically data gathering technologies such as MODE or the GREENGAGE app can be related to the DigiTwin. However, every gathered data set needs a spatial component to be mapped into the digital atlas. During the project this meant that only the data from the GREENGAGE app has been integrated.

Due to the focus on the Dutch context, the applicability of the DigiTwin in other GREENGAGE Pilots is limited. While its web-based nature makes it accessible, the data focus is on Dutch contexts, resulting in less shareable form in the GREENGAGE platform.

3.3.8 Multi-modal data processing tools

The following 3 tools have been added to GREEN Engine in iteration 2 in order to deal with processing of images from roads with defects and deal with text sentiment analysis and summarisation.

Dashcam-based Road Defects Detection (UWE): The image analysis work has been successfully extended to dashcam footage, aiming to spot a wide range of roadway defects for the Turano Pilot, including potholes, surface cracks, faded road markings, and vandalised road signs. As part of the initial training phase, we deployed CVAT24 to label representative frames from the dashcam videos. Further, we integrated Segment Anything Model (SAM) model inside CVAT to speed up and accelerate the annotation process. An AI model is currently under development, and an early prototype already demonstrates promising results, offering side-by-side visual analysis of identified defects. The next release will output structured JSON files detailing all roadway defects and respective metadata, enabling research and development for downstream urban use cases. This is an exciting and evolving area of work, with ongoing interactions to increase accuracy, usability, and scalability.

Sentiment analysis (UWE): Using Large Language Models (LLMs) - Flan-T5, sentiment analysis is performed on transcripts from videos, interviews, and written texts from citizens. Inputs are processed by the LLM to detect emotional tone and classified with a binary label (1 for positive sentiment and 0 for negative). This approach enables rapid assessment of overall emotional trends, helping to quickly understand the prevailing mood or attitude expressed across large volumes of collected data. The models used are run locally to avoid sharing information with third party software. Initial evaluation is compared against ground truth and results are promising.

Summarization and thematic analysis (UWE): As part of the Bristol Pilot, a text summarization service leveraging LLaMA 3 is being developed to summarize content from qualitative comments submitted by citizens through the EBLN25 platform. This tool is designed to help the Pilot owner and other stakeholders quickly grasp key themes from large volumes of textual survey data. To support

²⁴ <https://www.cvat.ai/>

²⁵ East Bristol Liveable Neighborhood

experimentation and broader usability, a streamlit web application has been created, enabling users to upload their own datasets and explore summarization across various contextual settings. Initial evaluations by domain experts have shown that summarisation quality can vary depending on the context provided, an insight that is guiding further refinement. Building on this foundation, the work is now being extended to include thematic summarization, aiming to group and highlight recurring topics and concerns more effectively. This evolving capability holds strong potential for enhancing civic engagement analysis and decision-making.

3.4 Horizontal services

The following sections detail some services that have been adopted in GREENGAGE which do not deal neither with the data management process nor with Citizen Science management process. Concretely, they tackle the need to have a centralized, cross-module, authentication and profile management service, implemented with Keycloak, and a chosen repository where the open datasets generated by the project will be made public.

3.4.1 Keycloak

Keycloak has been adopted as the identity and access management (IAM) system in GREENGAGE to offer a unified and secure way for users to authenticate across all digital services in the project. Rather than relying on separate login systems for each tool, Keycloak enables a smooth and consistent sign-in experience, making it easier for users to navigate the GREENGAGE ecosystem while keeping security and access control tightly managed.

Socio-demographic data

Keycloak also has been extended to support the storage of users' socio-demographic data by incorporating custom user attributes into its identity management framework. This modification allows the platform to capture and persist information such as age range, gender, education level, and occupation directly within each user's profile, using <https://me.greengage-project.eu> as frontend. These enhancements enable seamless integration with external applications requiring demographic segmentation or personalized user experiences, while maintaining compliance with privacy and security standards.

Please, select your pilot site or indicate your location, e.g. name of your place or ZIP code

Other ▼

Other

Bilbao

Gender

Male ▼

What is your level of education?

Master's degree ▼

What is your level of knowledge in the usage of digital tools?

Advanced level (I make purchases electronically, I use spreadsheets, ... ▼

Could you indicate your work status?

Employed (private sector) ▼

Please, select the type of organization you represent

Citizen ▼

Do you consider yourself to belong to a minority or socially disadvantaged group based around, for example, ethnicity, gender, economic or social background, or other?

No ▼

What is your age?

25-34 years – Early adulthood ▼

☒ I consent to taking part in the GREENGAGE research project based on the [Consent form conditions](#)

Figure 28: User view of the socio-demographic form.

Details
Attributes
Credentials
Role Mappings
Groups

Consents
Sessions
Identity Provider Links

Key	Value	Actions
userAttribute_ageRange	25-34	Delete
userAttribute_consent	2025-06-24T07:54:34.286Z	Delete
userAttribute_digitalToolsAcquaintance	Advanced level (I make pun	Delete
userAttribute_disadvantagedGroup	No	Delete
userAttribute_educationLevel	Master's degree	Delete
userAttribute_gender	Male	Delete
userAttribute_organizationType	Citizen	Delete
userAttribute_parentEmail		Delete
userAttribute_parentName		Delete
userAttribute_parentalConsent	-	Delete
userAttribute_schoolName		Delete
userAttribute_thematicRole	Core team member- organi	Delete
userAttribute_workStatus	Employed (private sector)	Delete
userAttribute_workStatus_other	false	Delete
userAttribute_zipCode	Bilbao	Delete
userAttribute_zipCode_other	Bilbao	Delete

Figure 29: Keycloak's administrator view of socio-demographic data.

Role within the Citizen Observatory

The context of GREENGAGE's Citizen Observatories, Keycloak plays a central role in managing who gets access to what, based on their role and the project area they're involved in. By supporting widely used standards like OpenID Connect and OAuth2.

Keycloak provides:

- **Dynamic role assignment**, enabling differentiated access for citizens, moderators, administrators, and developers within each observatory.
- **User group management linked to specific Pilots or campaigns**, allowing tailored access to missions, resources, and co-production processes.
- **Auditability and traceability**, logging all authenticated interactions to comply with the project's ethical and data protection frameworks.

Integration and Interoperability with Other Tools

Keycloak is fully integrated into the GREEN Engine infrastructure, acting as the unique identity provider for the following systems:

- **Collaborative Environment (CE)**: User identities and team contributions are linked to JWT tokens issued by Keycloak, enabling verified and trackable attribution of tasks and resources.
- **Discourse**: Operates with Keycloak-based SSO through OpenID Connect, automatically synchronizing user profiles and roles for community discussion and collaboration.
- **WordPress**: Utilizes compatible SSO plugins, ensuring users authenticated via the portal can navigate CE, Discourse, and other services without additional login steps.

- **GREENGAGE app and MindEarth for GREENAGE app:** Backend services consume Keycloak-issued tokens to authorize actions such as mission participation, data submission, and tracking.
- **Superset and data visualization tools:** Access is controlled via group and role-based permissions defined in Keycloak, ensuring users only access relevant datasets and dashboards.

Additionally, Keycloak is ready to support external login options, like university accounts or even social logins (e.g., Google, Apple), as long as these methods comply with GREENGAGE's privacy policies.

Ethical and Security Considerations

From the start, Keycloak was configured to comply fully with GDPR and the GREENGAGE Data Management Plan. Only minimal data (like email, user ID, and display name) is stored, and always encrypted.

On the security side, Keycloak allows administrators to define policies for token expiration, enable optional two-factor authentication, and control who has access to what through clear and strict permissions. Each service checks the token independently, ensuring tight control over access and reinforcing the principle of least privilege.

3.4.2 Zenodo

Zenodo²⁶ has been adopted in the GREENGAGE project as the primary open-access repository for hosting and disseminating datasets generated through the Thematic Co-Explorations carried out within the Citizen Observatories. By leveraging Zenodo's trusted infrastructure, supported by CERN and OpenAIRE, GREENGAGE ensures that its scientific outputs are publicly available, citable, and aligned with FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

Role within the Citizen Observatory

Within the context of GREENGAGE's Citizen Observatories, Zenodo plays a central role in the transparent publication and long-term preservation of data products resulting from citizen engagement, environmental sensing, and collaborative analysis. The dedicated GREENGAGE Zenodo Community serves as a centralized hub where datasets from different Pilots and thematic explorations in GREENGAGE can be accessed by researchers, policymakers, and the wider public.

Zenodo supports:

- **Dataset publication** with clear metadata and licensing, making it easier for others to discover, reuse, and cite the work.
- **Thematic organization**, grouping datasets by campaign, Pilot, or domain (e.g., air quality, biodiversity, noise pollution) for coherent access.
- **Version control and DOI assignment**, ensuring that each dataset is permanently identifiable and traceable to its source and version.
- **Integration and Interoperability with Other Tools**

Zenodo is integrated as part of the GREENGAGE data pipeline through automated or semi-automated data exports from the GREEN Engine and visualization dashboards. Data generated via apps, sensor platforms, or collaborative missions can be cleaned, curated, and then published to the Zenodo community. The metadata schema used in Zenodo aligns with the internal data cataloging standards, ensuring consistency between what is stored locally and what is shared publicly.

So far, the GREENGAGE Zenodo community hosts a small but rich collection of datasets and publications. The most comprehensive dataset comes from a Thematic Co-Exploration conducted at the University of Deusto's Bilbao campus, including anonymized participant demographics, survey responses for four campus locations, geolocated photos, air quality measurements from Atmotube Pro sensors, and pre/post impact assessment questionnaires—all structured in open formats. Additionally, the community features two academic publications: one detailing participatory design processes for

²⁶ <https://zenodo.org/>

visual analytics tools tailored to different stakeholder groups, and another focused-on co-production methodologies in Thematic Co-Explorations within Citizen Observatories. These resources collectively illustrate GREENGAGE's commitment to open data and participatory research practices.

Additional datasets will continue to be published as observatory campaigns progress, reinforcing GREENGAGE's commitment to open science and transparent data sharing.

Table 1: GREENGAGE components summary.

Area of Concern	Component	Description
Community and Co-production Process Management	Collaborative Environment	Platform for planning, managing, and tracking co-production processes in Citizen Observatories.
	Discourse	Discussion forum for collaborative communication, Q&A, and community interaction.
	WordPress	CMS providing informational landing pages and integration points for GREENGAGE tools.
Data Crowdsourcing and Curation	MindEarth for GREENGAGE app	Tool for crowdsourced geospatial image collection and mission-based mapping.
	GREENGAGE app	Citizen engagement app for mission-based data collection and interaction with policy makers.
	MODE	Smartphone-based tracking of mobility patterns for transport-related research.
	IoT Sensors	Low-cost sensors (e.g., Atmotube) for collecting air quality and noise data.
Analysis and Visualization for Insights Generation	Apache NiFi	Tool for data ingestion, transformation, and real-time processing.
	Apache Druid	Data store for fast, scalable, real-time analytics and time-series data.
	Apache Superset	Platform for creating dashboards and visualizing GREENGAGE data.
	Data Quality Dashboard	Interface for assessing and improving data quality and completeness.
	UrbanTEP / VISAT	Geospatial storytelling platform for sharing

		environmental indicators and analyses.
	DataHub	Metadata-rich repository and data management platform for GREENGAGE assets.
	DigitalTwin	3D visualization tool for urban planning insights and stakeholder dialogue.
	Multi-modal data processing tools	AI tools for road defect detection, sentiment analysis, and thematic text summarization.
Horizontal Services	Keycloak	Unified identity and access management system for all GREENGAGE tools.
	Zenodo	Open-access repository for publishing datasets and documentation from thematic explorations.

4 Usage manual of GREEN Engine and GREENGAGE's Academy

This section serves as a practical guide for stakeholders wishing to apply the GREENGAGE infrastructure and co-production enablers to execute a full Citizen Science (CS) Loop through Thematic Co-Explorations within Citizen Observatories (COs). Drawing on a validating experimental case study conducted at the University of Deusto campus in Bilbao, this manual illustrates the step-by-step usage of GREENGAGE, including its approach, GREEN Engine tools and enablers within its catalogue, for the transformation of citizen participation into actionable environmental intelligence.

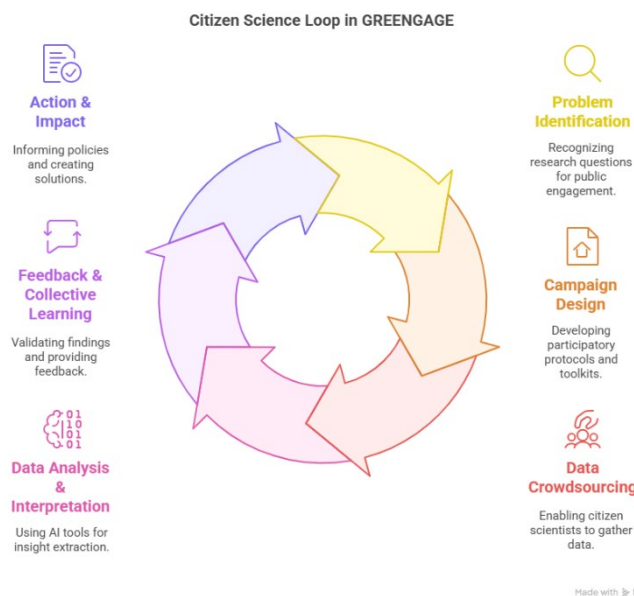


Figure 30: Citizen Science Loop in GREENGAGE, starting from “Problem Identification” and finishing in “Action&Impact”²⁷.

4.1 Overview of the Co-production Framework

GREENGAGE enables the structured execution of Thematic Co-Explorations through a four-phase co-creation process aligned with the six steps of the Citizen Science Loop (see Figure 30):

- **Phase 1: Preparing** → *Problem Identification*
- **Phase 2: Designing** → *Thematic co-exploration and Crowdsourcing Campaign design*
- **Phase 3: Experimenting** → *Data Crowdsourcing and Data Analysis & Interpretation*
- **Phase 4: Sharing** → *Feedback & Collective Learning and Action & Impact*

Each phase is operationalized via the [GREEN Engine](#)'s ecosystem of digital tools and knowledge assets, available at [GREENGAGE's Academy](#) and Collaborative Environment's [catalog](#). As described in the previous chapter, these are supported by digital identities, collaborative interfaces, and a modular backend capable of handling high-resolution environmental data.

²⁷ Generated with <https://www.napkin.ai/>

4.2 Setting up a Thematic Co-exploration

4.2.1 Preparing the Campaign in the Collaborative Environment

To initiate a Thematic Co-Exploration, coordinators must define a socio-environmental challenge. Using the "[Thematic Co-exploration Specification](#)" template, the following aspects are defined:

- **WHY** the Thematic Co-Exploration is relevant
- **WHO** the stakeholders and participants are
- **WHAT** activities will validate the hypothesis
- **WHEN** they will occur
- **WHERE** the geographical scope is
- **WHICH** tools and materials will be used
- **HOW** data analysis and storytelling will be carried out

In the validating exemplary Thematic Co-Exploration, which was organized on the campus of the University of Deusto in Bilbao, Spain, 10 researchers were recruited to evaluate four campus POIs (Points of Interest) based on air quality and spatial suitability. Each participant was trained in both the digital tools and the data collection protocol. Collaboratively they produced a full specification of the campaign to be realised.

Notably, they also worked on the definition of metrics to estimate air quality and campus space suitability. For that, two new metrics were co-defined by the team of observers in a meeting, namely, a **Perception of Air Quality Index (PAQI)** was made up, where on a scale from 1 to 5, people have to indicate their perception from very clean (no noticeable pollution effects) to highly polluted (major health concerns, unliveable conditions), and, the **Public Space Suitability Index (PSSI)** was also made up, where again in a 1 to 5 scale, volunteers have to express their perception regarding accessibility & connectivity (20%), safety & security (15%), environmental quality (15%), functionality & comfort (20%), sociability & inclusivity (15%) or maintenance & management (15%) aspects. Again, 5 ranges of suitability were defined ranging from excellent suitability (average answer values >4) to poor suitability (<1).

4.2.2 Stakeholder Onboarding with Discourse and Collaborative Environment

Participants registered via [GREENGAGE's identity manager](#) powered by Keycloak, provided anonymized socio-demographic data and signing consent forms. Keycloak's integration with the Collaborative Environment, WordPress, Discourse and all the other tools part of GREEN Engine, has ensured a smooth and unified sign-on experience. Communication was coordinated through the Discourse channel, enabling pre-campaign training, coordination, and post-campaign feedback. The creation of dedicated threads and polls helped engage participants continuously.

With the support of the Collaborative Environment a new co-production process to govern, support and trace the Thematic Co-Exploration for the campus in Bilbao of the University of Deusto was created. Notice that it follows the 4 phases indicated in 4.1. Notice that for each suggested task to be realized, enablers that may guide a team of observers in their collaboration are suggested and may be instantiated as shown in Figure 31.

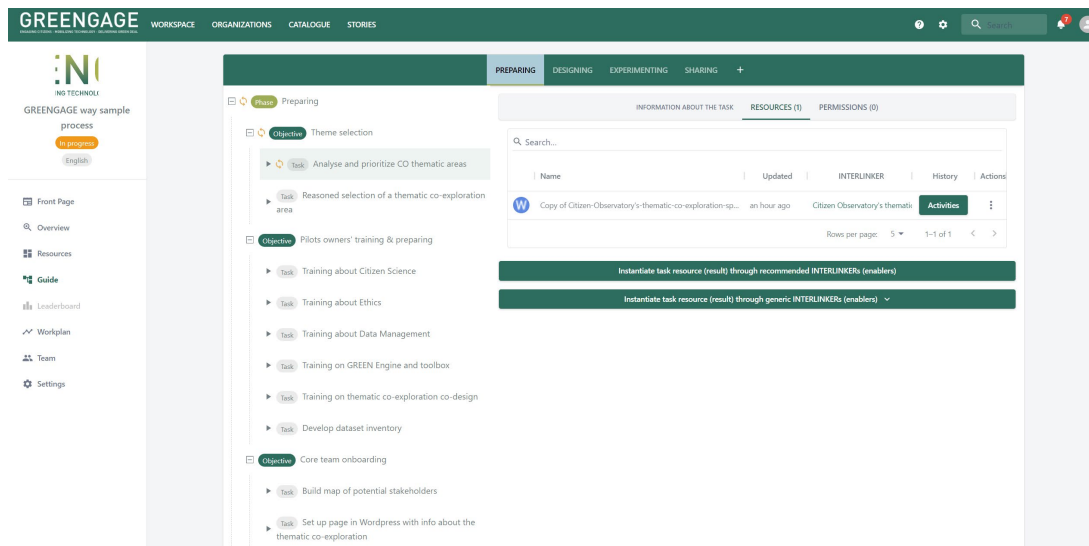


Figure 31: GREENGAGE Collaborative Environment

backend

4.3 Designing the Crowdsourcing Campaign

4.3.1 Configuring POIs and Tasks

Using the [GREENGAGE app's backend dashboard](#) (see Figure 32), whose fully usage details can be found at its [documentation](#), a new observatory instance is created. Each POI includes:

- **Location metadata** (type, description, GPS coordinates)
- **Tasks** such as:
 - Photo capture
 - Surveys on air quality (Perception of Air Quality Index, PAQI) and space suitability (Public Space Suitability Index, PSSI)

Task configuration is supported by a visual campaign builder interface. Survey templates are re-usable and customizable through a dedicated form builder. Metadata for tasks is recorded in the GREEN Engine's database for traceability and reuse.

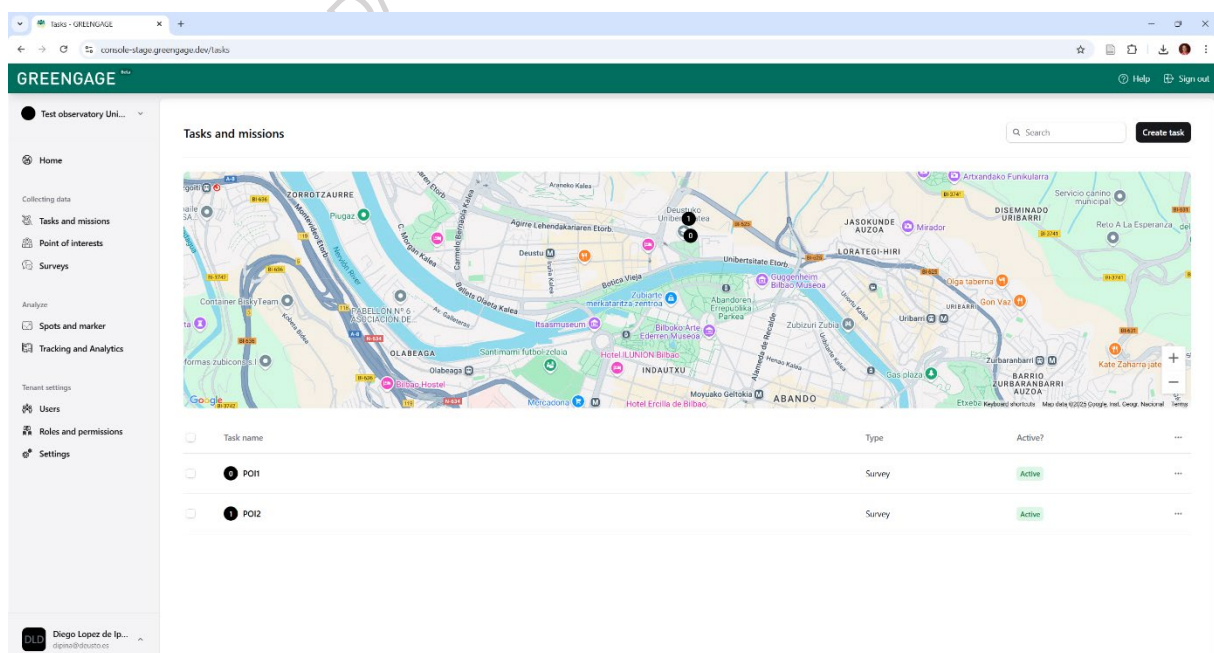


Figure 32: GREENGAGE app's backend dashboard

4.3.2 Deploying the GREENGAGE App

Volunteers download the app from [Google Play](#) or [Apple App Store](#). After login, participants (see Figure 33) are able to select a POI, complete surveys, and upload photos. The app provides real-time feedback upon data submission and notifies users when they have completed all assigned tasks. In the reported validating experimental case study, [Atmotube Pro sensors](#) were also used for air quality data logging during the exploration route.

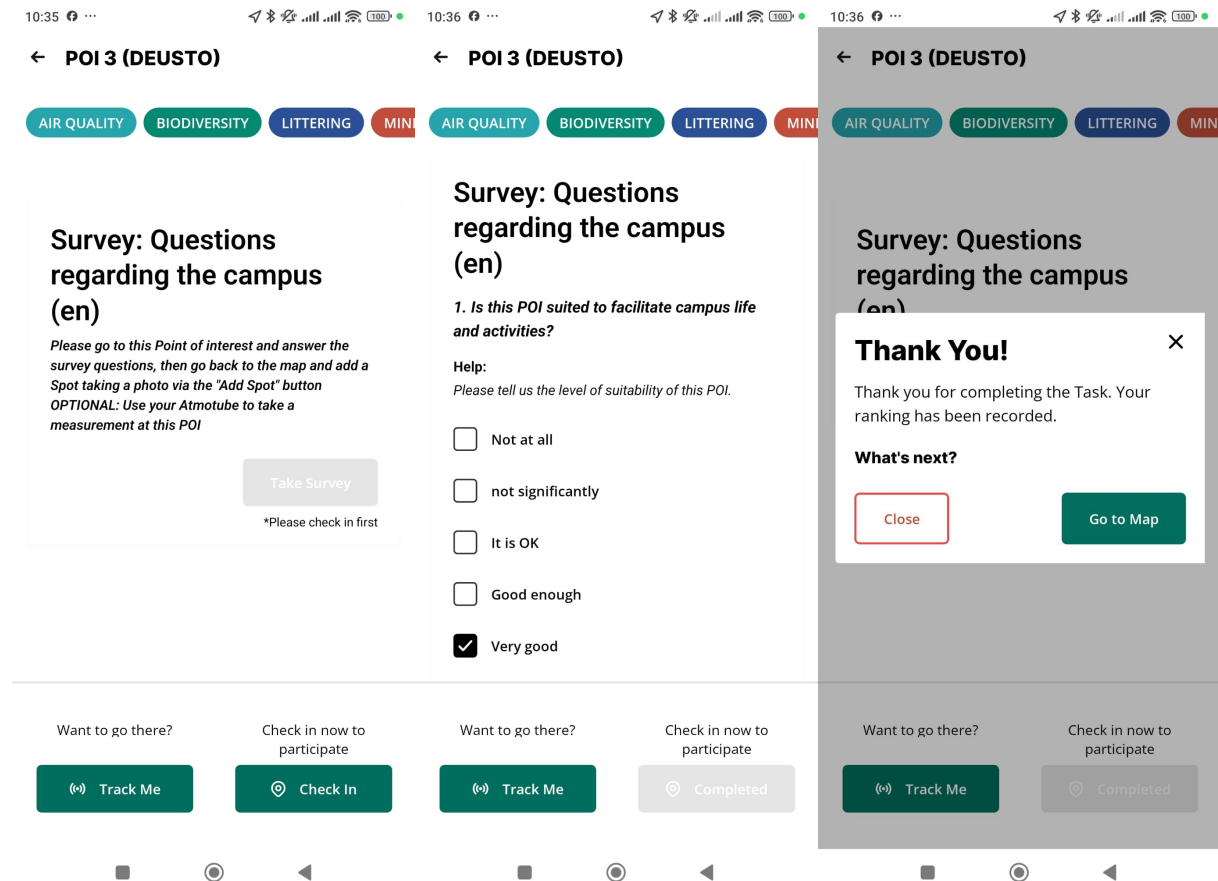


Figure 33: Three GREENGAGE app's screenshots used whilst observing POIs at Deusto University's campus.

4.4 Data Collection and Transformation

4.4.1 Executing the Campaign

On 14th March 2025, the campaign was conducted. Each of the 10 participants followed a guided journey via the GREENGAGE app, performing multiple tasks at four POIs:

- 21 photos were taken at 4 POIs
- 90 survey responses were submitted
- 180 air quality measurements collected by 4 Atmotube Pro devices

The campaign was scheduled in one hour and monitored through a real-time dashboard showing task completion status across POIs.

4.4.2 ETL and Data Integration

Data was extracted using [Apollo Server's](#) GraphQL API (see Figure 34). Two ETL²⁸ pipelines were run:

²⁸ ETL – data integration process, Extract (E), Transform (T), L (Load)

1. Photos and geodata
2. Survey responses

The ETL process was completed with some additional ad hoc Python scripts using asynchronous APIs to extract data from sociodemographic and PRE and POST impact questionnaires completed by participants in the campaign. Each record was enriched with user roles, timestamps, geolocation, and device metadata. Once transformed into structured CSVs, datasets were ingested into [Apache Druid](#) (see Figure 35). For traceability, each transformation step was logged, and reproducibility scripts were published as part of the GREENGAGE data pipeline repository.

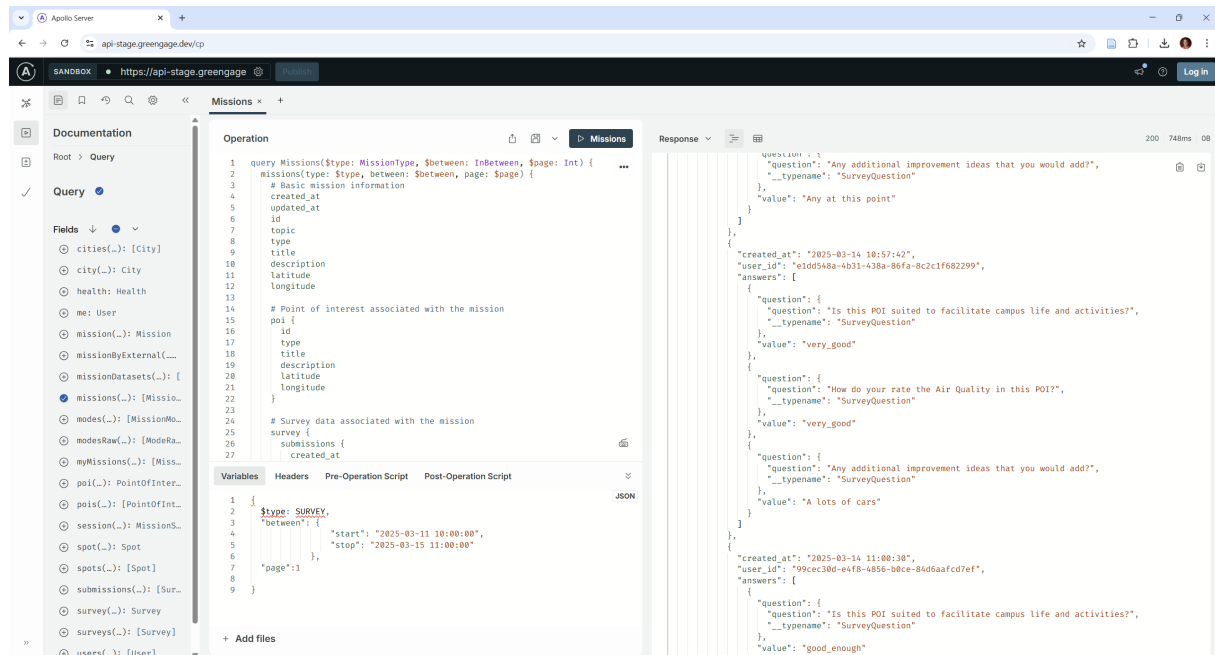


Figure 34: Apollo Server's GraphQL API.

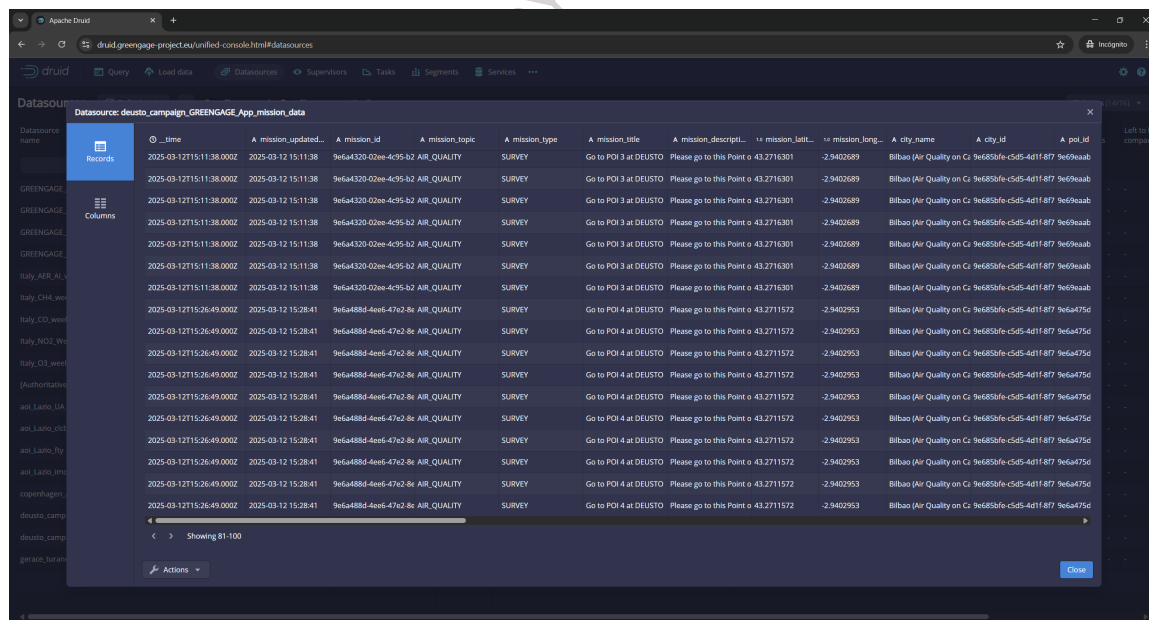


Figure 35: Apache Druid's data store for GREENGAGE.

4.5 Data Analysis and Visualization

4.5.1 Creating Dashboards with Apache Superset

Data stored in Druid was queried and visualized in [Apache Superset](#). Some of the charts produced are:

- Air quality perceptions vs. sensor readings (e.g., PM2.5) (see Figure 36)
- POI-specific and campus-wide summaries
- Photos depicting POIs status in campus (see Figure 37)

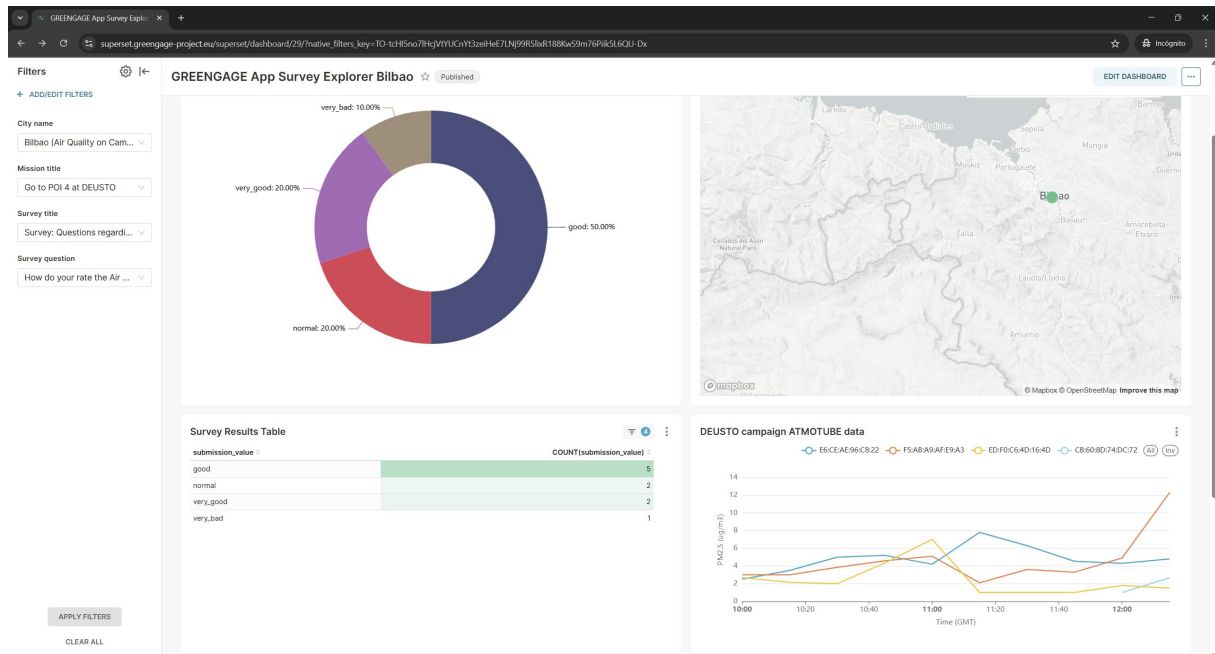


Figure 36: Per POI visualization of air quality perception and pollution detection via Atmotube.

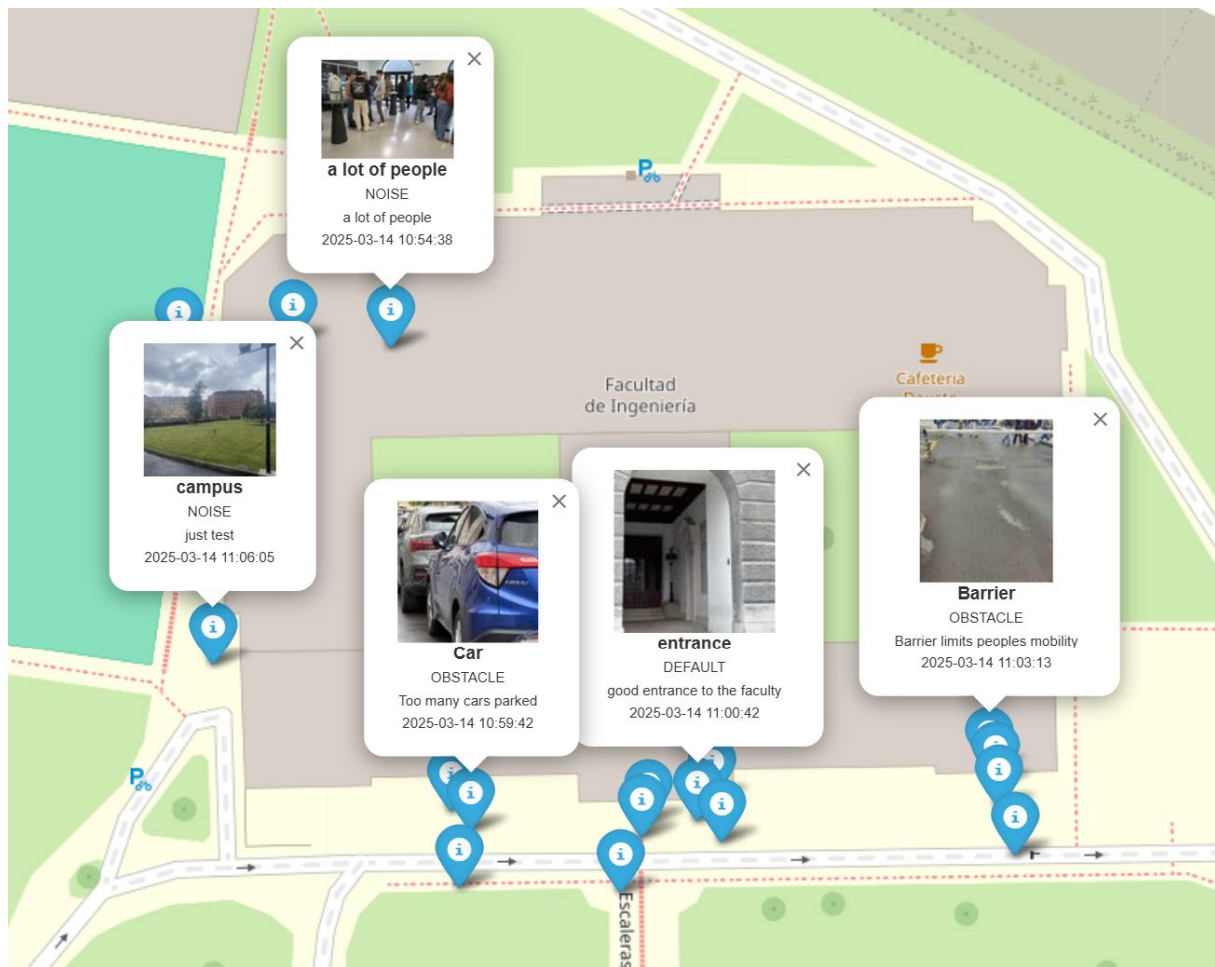


Figure 37: Ad-hoc visualization showing pictures taken by observers at different spots in Deusto's campus.

For instance, in our use case, POI 4 received a 70% "good" or "very good" rating from observers, while sensor data confirmed PM2.5 concentrations in a safe range (1–7.8 $\mu\text{g}/\text{m}^3$). Additional multi-view dashboards aggregated survey results by demographic group, supporting equity analysis.

4.5.2 Policy Dashboard and Reporting Templates

A tailored Superset dashboard was also prepared for policy stakeholders. This included simplified charts, PDF export capabilities, and metadata summaries. Visualizations were linked to recommendations and excerpts from participant feedback.

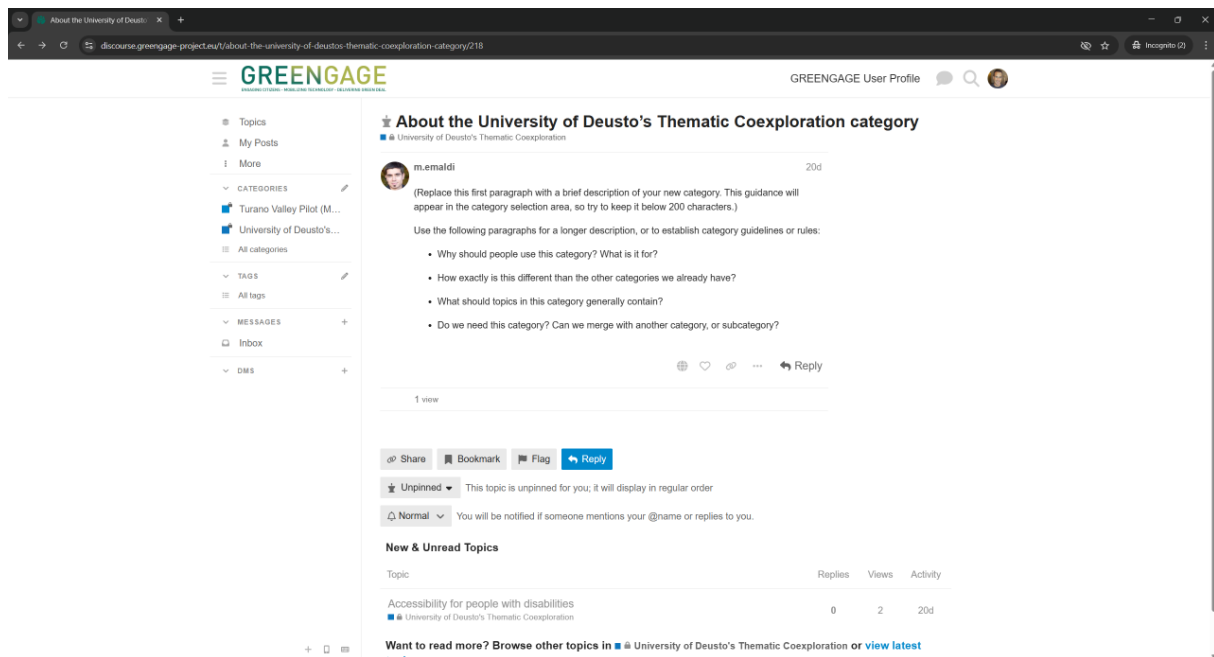


Figure 38: Discussion forum for Deusto's campus' thematic co-exploration.

4.6 Feedback, Impact and Dissemination

4.6.1 Policy and Social Engagement

The following actions were taken:

- Results were shared in the [Discourse](#) community (see Figure 38).
- Posts were disseminated via LinkedIn
- A policy brief was created following the [policy brief template](#) CO enabler and was then delivered to Deusto's Vice Chancellor. Participants reflected on results through guided discussion forums in Discourse. The feedback helped generate actionable insights, including campus planning suggestions and recommendations to enhance pedestrian safety near POIs (Figure 39).



Figure 39: Policy brief summarising recommendations for University of Deusto's campus' administrators.

4.6.2 Impact Assessment and Zenodo Publication

Impact was assessed using an adapted [ACTION project's impact evaluation approach](#), measuring social, scientific, and political relevance (see Figure 40). Datasets were published to [GREENGAGE's Zenodo community](#) (see Figure 41), including:

- Anonymised survey and demographic data
- Air quality sensor logs
- Photo datasets
- PRE and POST campaign evaluation responses

The uploaded datasets include clear metadata, licensing terms (CC-BY), and documentation to facilitate reuse. Notice that the impact assessment reported was generated following the ACTION's evaluation approach.

Impact Assessment Report – GREENGAGE Thematic Co-Exploration at University of Deusto's campus

1. Introduction

This report presents the impact assessment of the thematic co-exploration conducted as part of the GREENGAGE project at the University of Deusto's campus, performing a "reflection on the suitability and air quality of important points of interest (POIs) within the campus of the University of Deusto in Bilbao, Spain". This impact assessment evaluation has been carried out following [ACTION impact assessment framework](#). The experiment aimed to engage participants in a Citizen Science (CS) process to co-explore and reflect on issues related to environmental awareness, political participation, scientific engagement, and community empowerment.

A total of 10 participants took part in the pre- and post-evaluation surveys designed to capture changes in perceptions, behaviours, and engagement across several impact areas. As mentioned, this exemplary thematic co-exploration took place within the campus of the University of Deusto in Bilbao, SPAIN on Friday 14th March 2025, from 11:30am to 12:30pm CET

2. Methodology

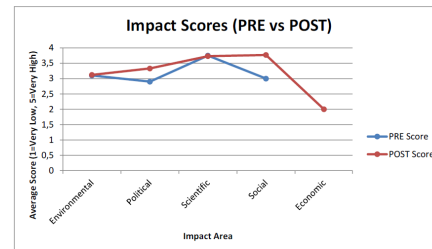
The ACTION framework outlines impact dimensions in five major areas: Scientific, Social, Political, Environmental, and Economic. Pre- and post-evaluation questionnaires were administered to the same group of participants to capture quantitative and qualitative changes. Quantitative scores were compared, and qualitative responses were analysed to assess transformation.

3. Summary of Quantitative Impact Comparison

The following table summarizes the average scores (\pm standard deviation) across impact dimensions from pre- and post-evaluation surveys:

Impact Area	PRE Score (AVG \pm SD)	POST Score (AVG \pm SD)	Change
Environmental	3.1 \pm 1.2	3.12 \pm 1.14	Slight decrease
Political	Qualitative: Low	3.33 \pm 1.26	Significant increase
Scientific	3.75 \pm 1.26	3.73 \pm 0.97	Slight stable
Social	Qualitative: Medium	3.77 \pm 1.32	Slight increase
Economic	Not measured	Qualitative: Low	Slight increase

Below, you can find a graphical representation of the impacts achieved at the thematic co-exploration held in the campus of the University of Deusto in Bilbao, SPAIN.



4. Key Impact Area Observations

4.1 Scientific Impact

Participants demonstrated high scientific literacy at the outset, with an average score of 4.55, which slightly increased to 4.58 in the POST experimentation's questionnaire. Many participants had existing research backgrounds. During the project, they reported increased involvement in Citizen Science campaigns, suggesting improved understanding and appreciation of science and CS tools. Participants had a positive perception towards Citizen Science which was kept varying from 3.75 to 3.73 from PRE to POST questionnaires' answers. Still, the standard deviation was reduced meaning that there was a more consistent positive view regarding Citizen Science by participants after concluding the thematic co-exploration. Anyhow, it must be admitted that there is clear scope for improvement regarding the "vision towards science" that participants in thematic co-explorations have. Such aspect was gathered by assessing participants' perceptions of science's benefits, its role in improving quality of life, its societal pace-setting influence, and its balance with non-scientific values like faith.

4.2 Environmental Impact

Environmental concern remained stable (slight decrease from 3.56 to 3.45). However, there was a clear improvement in ecological behaviours, as participants moved from 'often' to 'always' in adopting environmentally friendly practices. The shift from "Yes, often" to "Yes, always" across all 8 environmental behaviours indicates a strong reinforcement of pro-environmental habits.

Figure 40: Impact analysis performed following the ACTION project's impact evaluation approach.

Published April 11, 2025 | Version v1

Datasets collected from a GREENGAGE's thematic co-exploration for Bilbao campus at University of Deusto

López de Ipíña González de Artaza, Diego (Contact person)¹ ; Wolosiuk, Dawid (Data curator)²

The following datasets were collected in a thematic co-exploration realized at the campus in Bilbao of University of Deusto:

- Anonymized sociodemographic data of participants in the thematic co-exploration organized at University of Deusto's campus in Bilbao
- Anonymized dataset with aggregated survey answers for the 4 POIs defined at the campus of University of Deusto
- Dataset with all snapshots (photos) captured at different spots in the University of Deusto's campus
- Atmotube Pro sensors measurements during from 4 different devices (users) captured between 11:30 - 13:30 CET local time on 14th March 2025
- PRE (before the participation in the CS campaign) impact evaluation questionnaire answers and report
- POST (after the participation in the CS campaign) impact evaluation questionnaire answers and report.

Full details about the CS loop completed when executing this thematic co-exploration are available at: <https://github.com/Greengage-project/Documentation/blob/main/docs/CS-loop-GREENGAGE.md>

Files

Name	Size	Download all
deusto_atmotube_data_CB_60_8D_74_DC_72_20250318.csv	1.6 MB	
deusto_atmotube_data_CB_60_8D_74_DC_72_20250318.csv	1.0 kB	

25 VIEWS **49 DOWNLOADS**

Versions

Version	Date
Version v1	Apr 11, 2025

External resources

Indexed in

Communities

Figure 41: Datasets created in the Thematic Co-Exploration, uploaded into GREENGAGE's Zenodo community.

4.7 Conclusions and Recommendations

The GREENGAGE platform, through tools like the Collaborative Environment, Keycloak, GREENGAGE app, Apollo Server, Apache Druid and Superset and the suite of CO enablers part of GREENGAGE Academy, enables non-experts to effectively carry out a complete Citizen Science Loop. However,

expert facilitation remains essential in configuring ETL pipelines and interpreting complex analytics. Sustained support and feedback loops are key to fostering community engagement and scaling Thematic Co-Explorations across diverse territories.

Lessons learned include:

- Early training and role assignment improve campaign execution.
- Integrated dashboards improve transparency and stakeholder engagement.
- Publishing datasets and reports enhances credibility and replicability.

This usage manual (section 4) serves as an example for other Pilots or cities to replicate the Deusto campus experience, tailoring it to local environmental and community contexts while leveraging GREENGAGE's modular and interoperable toolset. Future deployments could explore integration with urban digital twins or extend participatory modelling approaches, thereby scaling the impact of GREENGAGE's Citizen Science methodology. A more detailed account about the whole process (step-by-step) carried out for the co-production of Deusto's campus Thematic Co-Exploration is available in GREENGAGE's official documentation page (Figure 42):

<https://greengage-project.github.io/Documentation/CS-loop-GREENGAGE/>

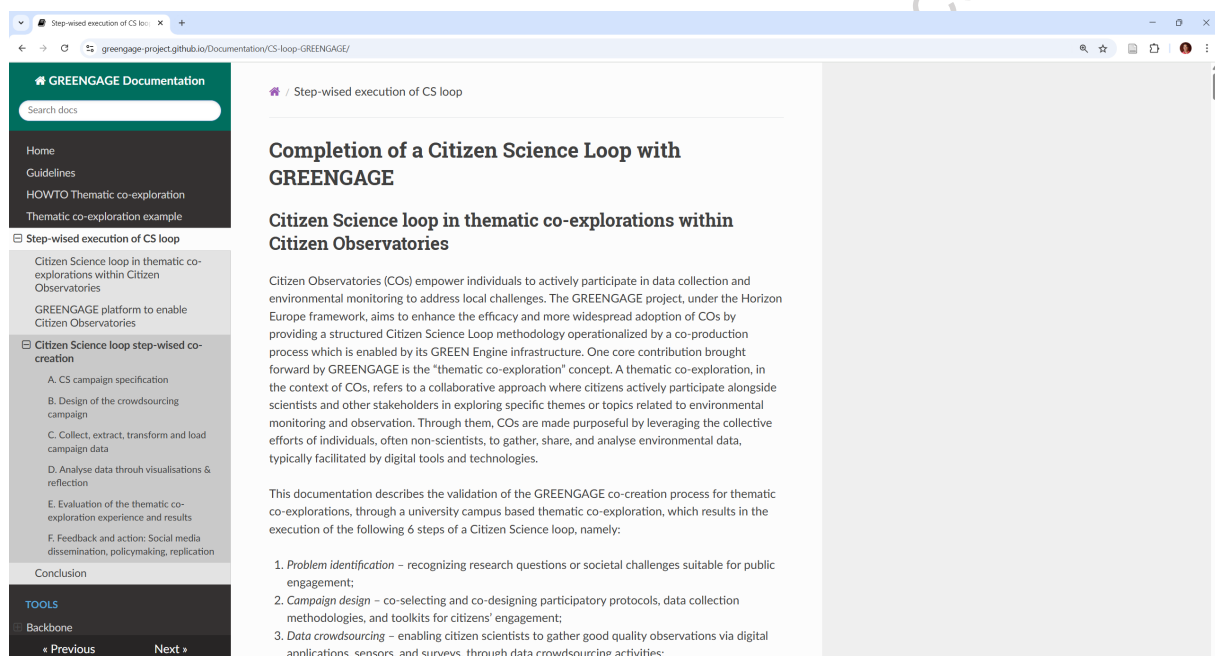


Figure 42: Technical documentation page where the step-wise co-production of a Thematic Co-Exploration following GREENGAGE's approach and using its tools and knowledge assets is detailed.

5 Usage statistics & validation

This section focuses on the evaluation and analysis of the impact these tools have on the project's overall success. The evaluation aids in understanding how each tool contributes to the project's goals and objectives. This analysis is not only crucial for current assessment but also lays the groundwork for developing future Key Performance Indicators (KPIs). These KPIs are instrumental in measuring and enhancing the project's performance, ensuring that the project's impact is both significant and quantifiable. This section reports the usage statistics until the end of June 2025, when the Pilots' execution is still ongoing.

5.1 Collaborative Environment

The data collected in the Collaborative Environment (CE) includes usage types that are extensive and cover various aspects relating to the co-production process. It comprises the management of co-production processes where events such as creating, reading, updating, or deleting a co-production process are defined and tracked through logs stored in a Elasticsearch²⁹ database. In addition to this, events have been defined to capture activities carried out on process settings. The other key focus area is management of organizations and teams whereby events are monitored for creation, updating and deletion of these entities. To complement these data collection methods based on logs management, the CE also uses Matomo tool. Matomo³⁰ is a widely used free open-source web analytics application which helps understand how users interact with the platform. It tracks online visits made to one or more websites and presents reports about them for analysis purposes. This enables us to know which pages are most often visited by whom from which location thereby providing useful information on user behavior and engagement with the CE. Hence, the CE is instrumented with logging and usage analytics instruments which help to extract usage statistics, resulting in a profound comprehension of how the platform works.

The CE within the project framework utilizes a series of Key Performance Indicators (KPIs) to effectively measure the achievement of goals by the Pilot. These specific KPIs encompass both global and local metrics, tailored to accurately reflect the platform's performance and efficacy. The KPIs are dynamic and will be adapted based on Pilot outcomes and lessons learned, ensuring they accurately represent the tool's functionality and success. To populate such KPIs exploration of Matomo analytics or queries of the logs gathered or data model used internally by the CE will be necessary.

To gather real-time log events (e.g., creation, updates, and deletions of co-production processes, teams, and organizations), we have leveraged the structured datasets stored in Collaborative Environment to analyze historical trends in usage patterns. The following visualizations (Figure 43, Figure 44 and Figure 4546) highlight key usage indicators extracted from the CE's internal database:

Coproduction Processes Created Per Month

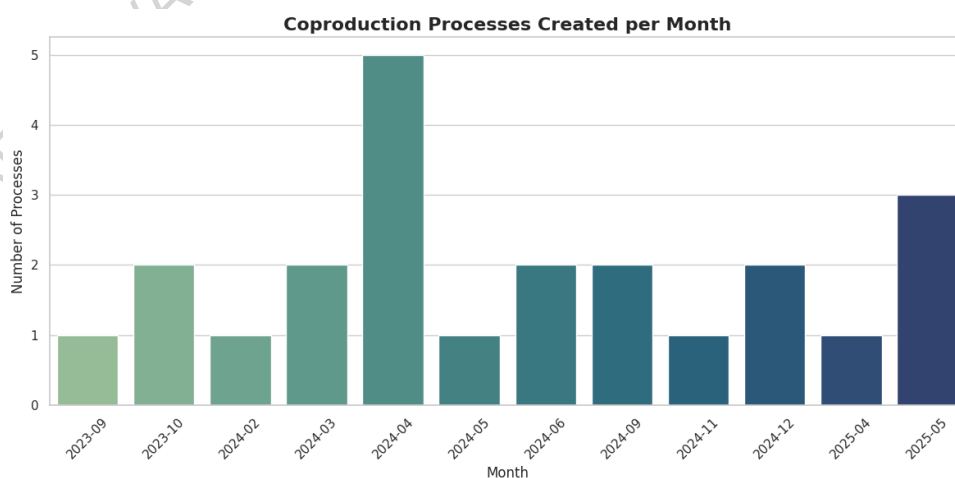


Figure 43: Chart depicting the processes created through Collaborative Environment from its beginning to June 2025.

²⁹ <https://www.elastic.co/>

³⁰ <https://matomo.org/>

The above figure displays the frequency of new co-production processes initiated by users. It reflects the evolution of collaborative project creation over time, serving as a proxy for platform engagement. Given that each Pilot is usually embarked in one or two Thematic Co-Exploration's per iteration and that we have 5 Pilots, the results are reasonable.

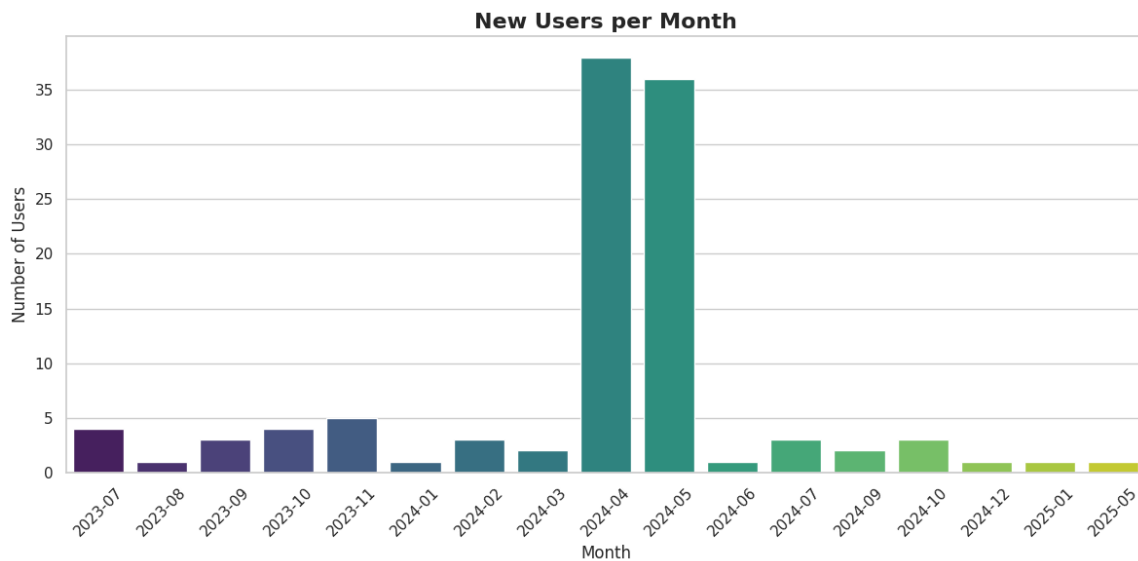


Figure 44: Chart depicting the users who have registered/entered for first time in the Collaborative Environment.

The above figure displays the number of new users per month. A peak occurred in May/June 2025 when Pilots' second iteration was launched.

Team Formation Activity

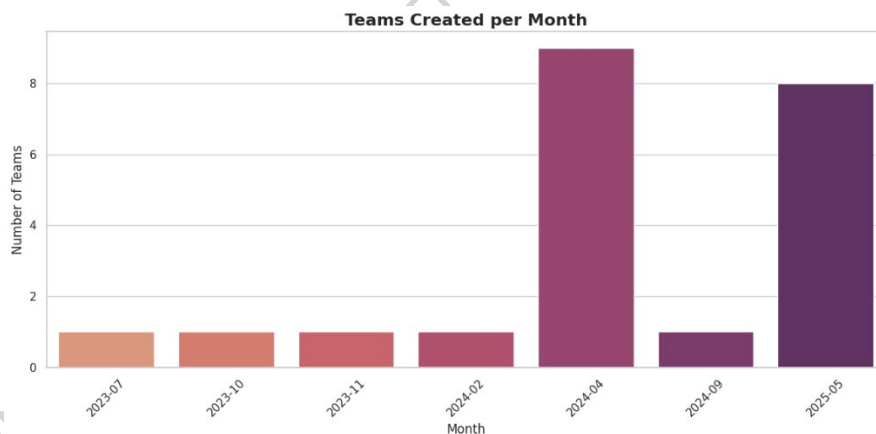


Figure 4546: Chart depicting the teams created in the Collaborative Environment.

Likewise, the above figure shows how new teams have been created along the project deployment.

These visual analytics not only support performance monitoring but also feed into the CE's KPI framework.

5.2 Discourse

Usage Statistics Extraction in Discourse provides various mechanisms for extracting usage statistics. One of the key tools is the Discourse Data Explorer plugin, which allows for detailed data analysis and reporting. This plugin can be utilized to create custom queries, enabling administrators to extract specific usage statistics that are relevant to their community's needs. The types of usage data that can be collected include metrics such as the number of active users, frequency of posts, and engagement levels. These metrics can be visualized and analyzed to gain insights into how users interact with the

forum. The Data Explorer plugin is a powerful tool for Discourse administrators to dive deep into the data and extract meaningful statistics for evaluation.

Methods Employed for Data Collection, Discourse forums can integrate with external analytics tools like Matomo to track user activity and engagement. This integration allows for a more comprehensive view of user behavior, including page views, session duration, and user demographics. Additionally, community feedback can be gathered through surveys or direct user interactions within the forum. This combination of in-built analytics and external tools provides a robust method for collecting a wide range of usage data.

Evaluating Tool Impact, the evaluation of Discourse's impact on Thematic Co-Explorations can be achieved by analyzing the extracted usage statistics. For instance, administrators can track trends in user engagement over time, identify the most active topics, and assess the growth of the community. By analyzing these metrics, it becomes possible to evaluate the effectiveness of Discourse in facilitating meaningful discussions and collaborations. Furthermore, this analysis can inform strategies to improve user engagement and participation, ultimately enhancing the overall impact of the tool on Thematic Co-Explorations.

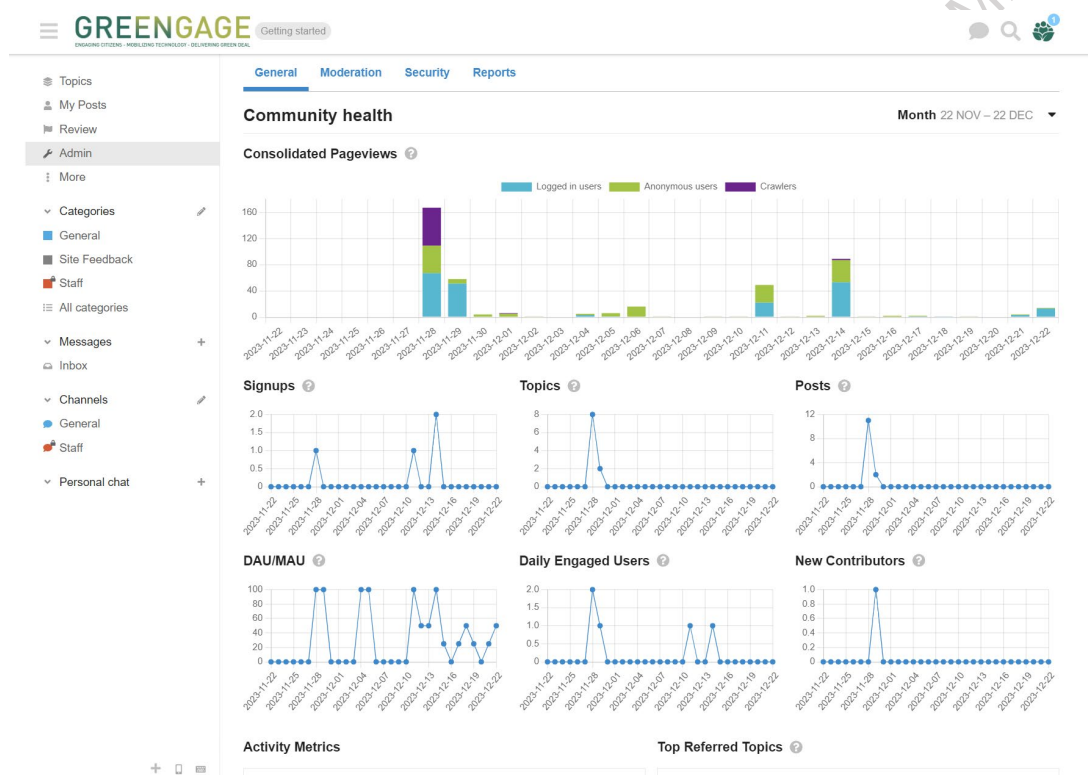


Figure 4748: An example of analytics built into Discourse deployment for GREENGAGE.

Since the inception of the project, the Discourse platform has generated a range of engagement statistics that reflect user activity and interaction (Figure 49). A total of 137 discussion topics and 165 individual posts have been created. Notably, a significant portion of this activity—84 posts—occurred during the alpha testing session of the GREENGAGE app and MindEarth for GREENGAGE app, which took place during the consortium meeting in North Brabant in April 2024. The platform has recorded a cumulative total of 1,206 pageviews. The average response time to user posts has been measured at approximately 26 hours, indicating a moderate level of responsiveness within the community.



Figure 49: Different statistics gathered at Discourse. Topics created (upper left), posts created (upper right), pageviews (bottom left) and time to first response (bottom right).

Figure 49: Different statistics gathered at Discourse. Topics created (upper left), posts created (upper right), pageviews (bottom left) and time to first response (bottom right).

Figure 49: Different statistics gathered at Discourse. Topics created (upper left), posts created (upper right), pageviews (bottom left) and time to first response (bottom right).

- Figure 49: Different statistics gathered at Discourse. Topics created (upper left), posts created (upper right), pageviews (bottom left) and time to first response (bottom right).

- Interpretation involves analysing these data points to understand user behavior and engagement.

Regarding “**Evaluation of Tool Impact**”, by analyzing the usage statistics from Matomo, you can evaluate the impact of the Citizen Observatory in several ways:

- *Understanding User Engagement*: Determine which parts of the site or which tools are most engaging or underused.
- *Audience Analysis*: Identify the demographic and geographic profile of your users to tailor content more effectively.
- *Content Effectiveness*: Understand what content is most appealing or educational for users.
- *Feedback Incorporation*: Use survey data to make informed decisions about new features or content.

Through Matomo, some of the evaluation questions that could be answered are:

- How many users are actively engaging with the observatory's tools on a monthly basis?
- What are the peak times for user activity on the site?
- Which geographical regions show the most interest in the observatory?
- How do new features or content affect user engagement?
- Are there any patterns in the way users navigate through the site?
- What are the common exit points on the website, indicating possible areas of user disinterest or confusion?
- How effective are external marketing and outreach efforts in bringing new users to the site?

By using Matomo with WordPress for a Citizen Observatory, you can gain insightful data that not only helps in understanding user behavior but also in shaping the future strategy and content of the observatory, making it more effective and impactful.

Based on data collected from the GREENGAGE website, the following insights can be noted. As of the time of writing this deliverable, the site has attracted approximately 7,500 unique visitors and recorded over 30,000 pageviews. The majority of visits have originated from Spain, the UK, Germany, the USA, and Denmark. Aside from the homepage, the most frequently visited sections are the GREENGAGE Academy and the Citizen Observatories, accounting for 15% and 11% of total visits, respectively.

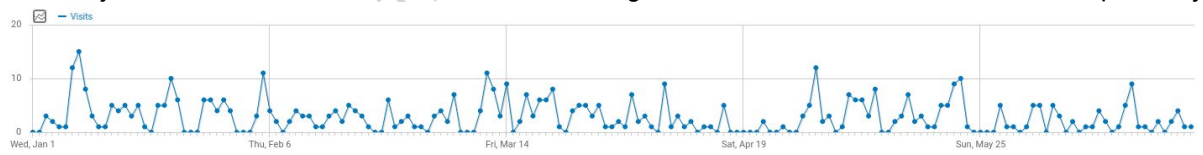


Figure 50: overview of the statistics from the current year.

In terms of external referrals, Google has been the primary source of traffic to the website, followed by Bing and Yandex. Among social media platforms, LinkedIn has generated the highest number of referrals, followed by YouTube, Facebook, and GitHub. However, most of the visits (75%) have been direct entries to the website.

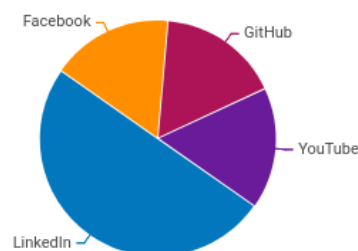


Figure 51: Distribution of visits from social networks.

5.4 MindEarth for GREENGAGE app

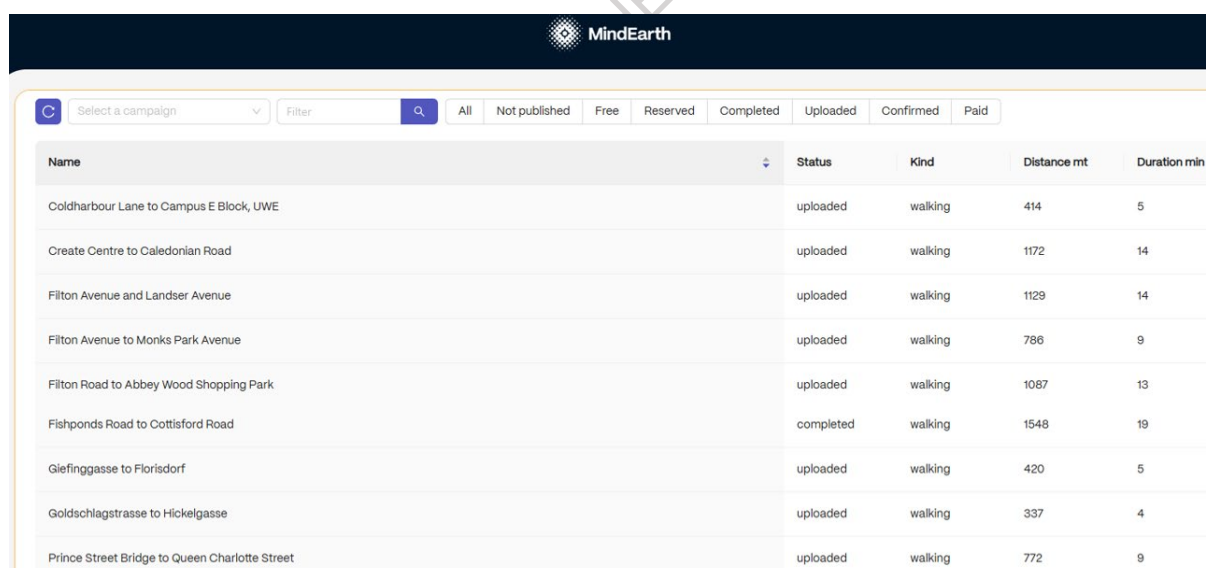
So far, test missions have been deployed across different Pilot to assess the usability and data collection capacity of the application. In a later stage, MindEarth for GREENGAGE app is expected to be more extensively used in Bristol and Copenhagen Pilots. In Bristol, while test missions have already been provided, the specific detection targets or objects have not yet been clearly defined. In contrast, in the Copenhagen Pilot, the app is expected to be used for vehicle detection tasks, with a particular focus on counting the number of trucks circulating in specific urban areas.

Usage statistics of MindEarth for GREENGAGE app is related to the number of users registered in both Android and iOS version of the app. So far, the platform registers 46 users, of which 7 are testers of the iOS version of the app. For the testing phase of the app, 13 campaigns have been activated across different Pilots (Table 2).

Table 2: Summary of published campaigns for testing.

Name of campaign	Number of missions available	Number of missions completed
Copenhagen Traffic and Air Quality monitoring – Area 01 - 10	228	2
Radstadt - Austria	2	1
Vienna Test Campaign 2025	3	2
Bristol Test Campaign 2025	11	10

Additionally, in Figure 52 is possible to have an overview of several missions' status on Mission Control platform, including distance walked and time spent. On next steps, further campaigns should be integrated into the platform in order to detect objects defined in Bristol and Copenhagen Pilots.



MindEarth					
Select a campaign		Filter	Q	All	Not published
				Free	Reserved
				Completed	Uploaded
				Confirmed	Paid
Name	Status	Kind	Distance mt	Duration min	
Coldharbour Lane to Campus E Block, UWE	uploaded	walking	414	5	
Create Centre to Caledonian Road	uploaded	walking	1172	14	
Filton Avenue and Landsers Avenue	uploaded	walking	1129	14	
Filton Avenue to Monks Park Avenue	uploaded	walking	786	9	
Filton Road to Abbey Wood Shopping Park	uploaded	walking	1087	13	
Fishponds Road to Cottisford Road	completed	walking	1548	19	
Gleifingasse to Florisdorf	uploaded	walking	420	5	
Goldschlagstrasse to Hickelgasse	uploaded	walking	337	4	
Prince Street Bridge to Queen Charlotte Street	uploaded	walking	772	9	

Figure 52: Mission control records for testing missions.

5.5 GREENGAGE app

Multiple Pilot cities have applied the GREENGAGE app to their research use cases within the duration of the project. The app has found most use with Gerace/Turano communities as well as North Brabant Pilot. There's furthermore an active discourse about the application and available extensions for Bristol and Copenhagen. Focal breakthrough lies on the availability of mission configuration in the backend for the Pilot Owners with the assistance of project coordinators and the dev team responsible for the App and backend support. Final goal is to reach the point of seamless and standardized process, where the Pilots would be entirely self-sustainable in conducting their desired research. The scope of applications

is different across the three Pilot regions. Gerace/Turano communities focus most extensively on the perceived air quality and greenery, North Brabant is applying the surveys method together with photo reporting to identify bike route issues, whereas Copenhagen aims to measure noise pollution and traffic density. Bristol will measure the perceived impacts experienced by residents after the installation of traffic modal filters and other street interventions as part of a Liveable Neighborhood scheme by the local authority.

Usage statistics can be separated into direct App Users and the backend Dashboard users. So far, the app has been downloaded 48 times on Android devices and 77 downloads on IOS devices.

Extensive use and testing of the apps reveal impressive results. Since the rollout of version 2 alone, a total of 27 observatories, 47 tasks, 24 surveys (with approximately 131 submissions), 94 spots, and nearly 350 MODE sessions have been recorded. Combined with version 1 (tracked separately), over 100 missions have been completed in total, generating around 0.75 million API requests during usage. In total more than 100 active users have been counted within the GREENGAGE app ecosystem.

Additionally, thanks to image optimization, version 2 has significantly reduced resource usage—saving approximately 1 GB in image traffic and around 500 MB in storage (excluding MODE files).

5.6 MODE

MODE itself does not collect any statistics; it is a service which can be used to extract trips from smartphone movements. However, the host application can store these trips to collect meaningful statistics:

- **Types of data:** Number of trips, modal split, duration and distance of trips.
- **Methods employed for data collection:** A database on the host app's side to store the trips.
- **The process for accessing and interpreting these statistics:** To be defined by the host app.

In a before/after study, a comparable set of trips shall be compared to see if any proposed changes to the infrastructure influenced the modal split, but also other KPIs like average trip lengths.

During the testing phases of versions 1 and 2, nearly 380 MODE sessions were processed by the backend infrastructure—approximately 350 of them during version 2 alone.

5.7 IoT Hyperlocal Sensors

Regarding the IoT sensors and environmental data collection, HOPU is leading on tasks related to air quality metrics that could be integrated into the GREENGAGE Citizen Observatory e.g., air quality index, mean and variance of each one of the pollutants collected by the sensors or perhaps the most polluted places in a city. In addition, MOONDIAL conducted wearable technology off-the-shelf research that validated the idea to use of the Atmotube for various partners. Additionally, HOPU is currently examining and exploring the use of the API Cloud and other services for the integration into the GREENGAGE backend.

The integration of the data collected by Atmotube Pro and all the metrics relating with the wearable sensors based on the location will be implemented with the Atmotube API Cloud. This service is provided by Atmotube company, is free to use and a simple description of its architecture is shown in Figure 53.

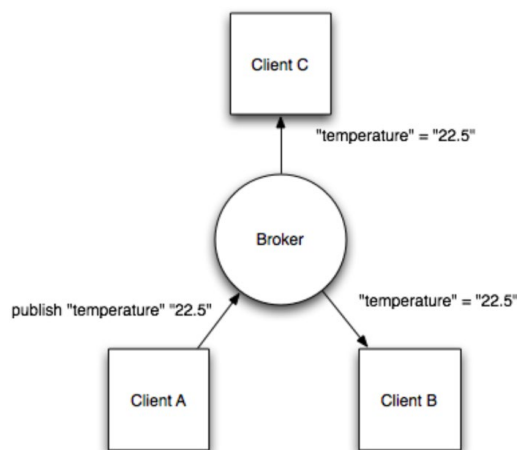


Figure 53: Example of interactions between MQTT Clients through an MQTT Broker.

Atmotube devices have been used in Denmark's Pilot where 17 Devices have been used and 65426 records gathered. In Italy, 20 Devices were used to gather 1950336 records.

5.8 Apache NiFi

Apache NiFi was finally not used in the project; instead, the data-acquisition team implemented 3 bespoke, programmatic pipelines to retrieve and ingest our sensor and environmental data. The first pipeline polls the Atmotube API to collect particulate and gas sensor readings; the second periodically fetches meteorological measurements from regional weather stations; and the third aggregates air-quality indicators from municipal monitoring networks. Each pipeline applies necessary validation and transformation logic in code before writing its output directly into dedicated Apache Druid data sources. This approach leveraged the team's existing development workflows and provided fine-grained control over error handling and scheduling.

5.9 Apache Druid

The Apache Druid cluster powering the data-ingestion pipeline has been instrumented to capture comprehensive usage and storage metrics. As of the latest measurement, the cluster hosts 32 distinct data sources, each representing either an individual device stream or a logical grouping of sensors. Across all data sources, a cumulative total of 7.681.231 rows has been ingested and retained, occupying 157,88 MB of on-disk storage. This corresponds to an average per-row footprint of approximately 21 bytes, underscoring Druid's efficient columnar compression and bitmap indexing strategies.

Datasources Refresh Show unused Show segment timeline									
Datasource name	Availability	Historical load/drop queues	Total data size	Running tasks	Segment rows minimum / average / maximum	Total rows	Avg. row size (bytes)	Replicated size	
ATMOTUBE_Air_Qual...	Fully available (9 segments)	No segments to load/drop	794.05 KB	No running tasks	0.001 M 0.007 M 0.008 M	65,426	12	794.05 KB	
ATMOTUBE_Air_Qual...	Fully available (9 segments)	No segments to load/drop	1.10 MB	No running tasks	0 0.007 M 0.027 M	65,426	16	1.10 MB	
ATMOTUBE_Air_Qual...	Fully available (108 segments)	No segments to load/drop	18.33 MB	No running tasks	28 0.016 M 1.555 M	1,684,285	10	18.33 MB	
ATMOTUBE_Air_Qual...	Fully available (66 segments)	No segments to load/drop	24.62 MB	No running tasks	4 0.027 M 0.174 M	1,812,180	13	24.62 MB	
Air Quality on Roads ...	Fully available (1 segment)	No segments to load/drop	12.19 MB	No running tasks	0.061 M 0.061 M 0.061 M	61,177	199	12.19 MB	
BRISTOL_Stop_and_s...	Fully available (396 segments)	No segments to load/drop	2.78 MB	No running tasks	1 18 39	7,201	386	2.78 MB	
BRISTOL_accidents	Fully available (1,577 segments)	No segments to load/drop	10.23 MB	No running tasks	1 3 12	4,265	2,398	10.23 MB	
BRISTOL_collisions	Fully available (181 segments)	No segments to load/drop	10.72 MB	No running tasks	141 272 412	49,316	217	10.72 MB	
BRISTO_street_level...	Fully available (15 segments)	No segments to load/drop	29.35 MB	No running tasks	0.013 M 0.014 M 0.016 M	212,290	138	29.35 MB	
Bristol_Air_Quality_C...	Fully available (30 segments)	No segments to load/drop	26.76 MB	No running tasks	0.009 M 0.051 M 0.096 M	1,537,945	17	26.76 MB	
Bristol_Redfield_Sch...	Fully available (1 segment)	No segments to load/drop	10.32 KB	No running tasks	153 153 153	153	67	10.32 KB	
GREENGAGE_APP_Mi...	Fully available (2 segments)	No segments to load/drop	265.45 KB	No running tasks	0 30 59	59	4,499	265.45 KB	
GREENGAGE_APP_Mi...	Fully available (1 segment)	No segments to load/drop	55.50 KB	No running tasks	470 470 470	470	118	55.50 KB	
GREENGAGE_APP_Mi...	Fully available (1 segment)	No segments to load/drop	40.11 KB	No running tasks	351 351 351	351	114	40.11 KB	
GREENGAGE_APP_Mi...	Fully available (1 segment)	No segments to load/drop	16.05 KB	No running tasks	28 28 28	28	573	16.05 KB	
GREENGAGE_APP_M...	Fully available (1 segment)	No segments to load/drop	355.55 KB	No running tasks	85 85 85	85	4,182	355.55 KB	
GREENGAGE_APP_Te...	Fully available (1 segment)	No segments to load/drop	57.43 KB	No running tasks	200 200 200	200	287	57.43 KB	
Italy_AER_Ai_weekly...	Fully available (1 segment)	No segments to load/drop	1.65 MB	No running tasks	0.132 M 0.132 M 0.132 M	131,922	12	1.65 MB	
Italy_CH4_weekly_20...	Fully available (1 segment)	No segments to load/drop	873.98 KB	No running tasks	0.118 M 0.118 M 0.118 M	117,936	7	873.98 KB	
Italy_CO_weekly_201...	Fully available (1 segment)	No segments to load/drop	1.65 MB	No running tasks	0.132 M 0.132 M 0.132 M	131,922	12	1.65 MB	
Italy_NO2_Weekly_20...	Fully available (1 segment)	No segments to load/drop	1.65 MB	No running tasks	0.132 M 0.132 M 0.132 M	131,544	12	1.65 MB	
Italy_O3_weekly_201...	Fully available (1 segment)	No segments to load/drop	566.03 KB	No running tasks	0.128 M 0.128 M 0.128 M	127,764	4	566.03 KB	
[AuthoritativeData] A...	Fully available (1 segment)	No segments to load/drop	393.88 KB	No running tasks	0.040 M 0.040 M 0.040 M	39,708	9	393.88 KB	
aoi_Lazio_CLC	Fully available (1 segment)	No segments to load/drop	419.59 KB	No running tasks	378 378 378	378	1,110	419.59 KB	
aoi_Lazio_UA	Fully available (1 segment)	No segments to load/drop	257.80 KB	No running tasks	378 378 378	378	681	257.80 KB	
aoi_Lazio_cicbb	Fully available (1 segment)	No segments to load/drop	127.68 KB	No running tasks	378 378 378	378	337	127.68 KB	
aoi_Lazio_ftv	Fully available (1 segment)	No segments to load/drop	79.74 KB	No running tasks	378 378 378	378	210	79.74 KB	

Figure 54: Dashboard View of Apache Druid.

Ingestion is orchestrated by six supervisor tasks, each scheduled to run every four hours (six times per day). These tasks fetch, parse, and index the latest sensor readings from the Atmotube fleet. The consistent recurrence of these workflows ensures that downstream analytics and user-facing dashboards always query against fresh data.

From the Data sources View (see Figure 54), we observe a uniform segment distribution, with each data source comprising between 10 and 25 segments. Individual data source sizes vary from approximately 3 MB to 8 MB, reflecting heterogeneity in device activity levels.

Task ID	Group ID	Type	Datasource	Status	Created time	Duration
index_parallel_ATMOTUBE_Air_Quality_Gerace_Turano_knjikiec_2025...	index_parallel_ATMOTUBE_Air_Quality_Gerac...	Index_parallel	ATMOTUBE_Air_Quality_Gera...	SUCCESS	2025-05-20T01:00:13.455Z	0:00:11
index_parallel_ATMOTUBE_Air_Quality_Gerace_Turano_daffigcc_2025...	index_parallel_ATMOTUBE_Air_Quality_Gerac...	Index_parallel	ATMOTUBE_Air_Quality_Gera...	SUCCESS	2025-05-20T01:00:09.984Z	0:00:14
index_parallel_ATMOTUBE_Air_Quality_Gerace_Turano_appfdiol_202...	index_parallel_ATMOTUBE_Air_Quality_Gerac...	Index_parallel	ATMOTUBE_Air_Quality_Gera...	SUCCESS	2025-05-20T01:00:06.555Z	0:00:12
index_parallel_gerace_turano_ATMO_gmcabbbi_2025-05-20T00:00:15...	index_parallel_gerace_turano_ATMO_gmcabab...	Index_parallel	gerace_turano_ATMO	SUCCESS	2025-05-20T00:00:15.622Z	0:00:10
index_parallel_gerace_turano_ATMO_kobeegh_2025-05-20T00:00:12...	index_parallel_gerace_turano_ATMO_kobeegh...	Index_parallel	gerace_turano_ATMO	SUCCESS	2025-05-20T00:00:12.059Z	0:00:11
index_parallel_gerace_turano_ATMO_aehepnme_2025-05-20T00:00:0...	index_parallel_gerace_turano_ATMO_aehepn...	Index_parallel	gerace_turano_ATMO	SUCCESS	2025-05-20T00:00:08.984Z	0:00:11

Figure 55: Task View of Apache Druid.

The Tasks View (see Figure 55) further validates operational stability: all supervisors report successful runs. These tasks are run at midnight to avoid possible Apache Druid usage peak hours. No ingestion failures have been recorded in the reporting period, indicating robust error-handling.

5.10 Apache Superset

Apache Superset is an open-source business intelligence (BI) tool designed to simplify data visualization and exploration. Developed by Airbnb and maintained by the Apache Software Foundation, it enables organizations to create dynamic, interactive dashboards and charts, offering intuitive ways to derive insights from vast amounts of data.

In our current project deployment, Apache Superset has played a crucial role by supporting efficient data visualization and informed decision-making. Through its intuitive user interface, team members can create diverse charts, dashboards, and datasets, enhancing collaboration and streamlining data-driven workflows.

As of the most recent review, our Superset instance contains a substantial number of data assets and visualizations. Specifically, the platform hosts **121 charts, 41 datasets, and 14 dashboards**. Among these dashboards, 12 have been published, indicating active use and dissemination within our team or broader organizational audience. This high rate of published dashboards suggests strong engagement and relevance of the visual insights being shared.

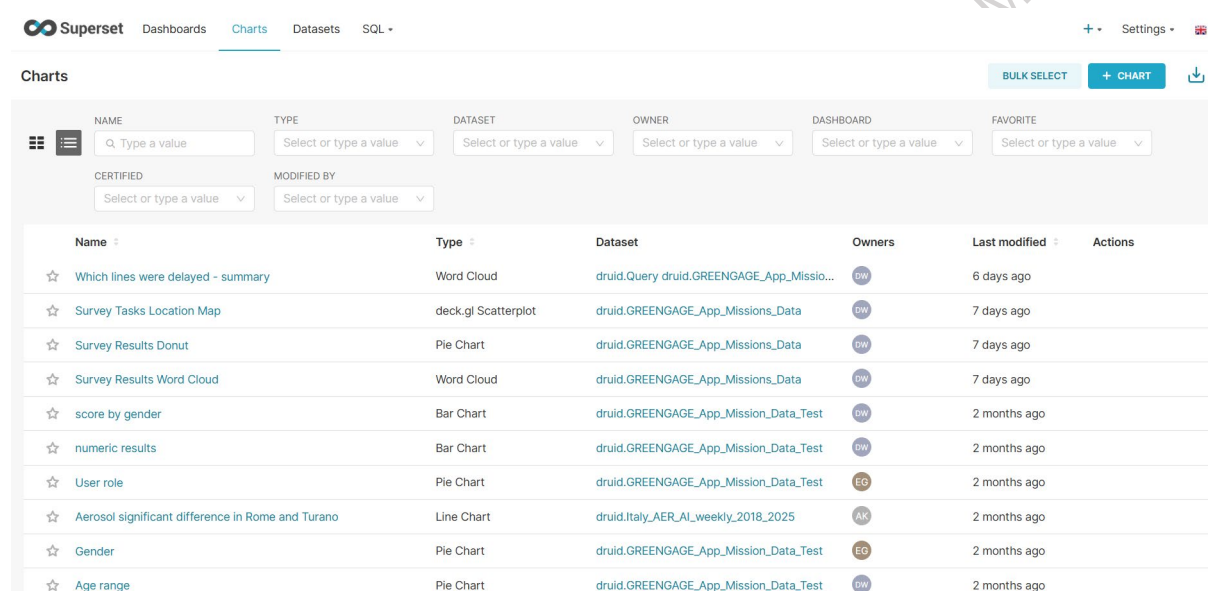


Figure 56: Superset Chart view.

The Charts screen in Figure 56 highlights a range of visualizations like Word Clouds, Bar Charts, Pie Charts, Scatterplots, and Line Charts. Diverse visualization types suggest varied analytical use cases within the project, with charts covering data such as survey results, demographic distributions, and geographical data analyses. Frequent updates, noted from recent modifications, reflect ongoing analytical activities.

Meanwhile the dataset screen in Figure 57 shows recent updates that indicate ongoing data integration, with the latest dataset modifications occurring as recently as 18 hours ago. This continuous dataset management underscores active data exploration and the maintenance of up-to-date information sources.

Superset

Dashboards

Charts

Datasets

SQL

+ -

Settings

Datasets

BULK SELECT

+ DATASET

NAME

TYPE

DATABASE

SCHEMA

OWNER

CERTIFIED

Q Type a value

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Select or type a value

MODIFIED BY

Select or type a value

Name	Type	Database	Schema	Owners	Last modified	Actions
<div></div> Query druid.GREENGAGE_App_Missions_Data	Virtual	Other	druid	<div></div>	18 hours ago	
<div></div> Untitled Query 1	Virtual	Other	druid	<div></div>	6 days ago	
<div></div> GREENGAGE_App_Missions_Data	Physical	Other	druid	<div></div>	6 days ago	
<div></div> GREENGAGE_App_Mission_Data_Test_socio	Physical	Other	druid	<div></div>	9 days ago	
<div></div> aol_Lazio_Imd	Physical	Other	druid	<div></div>	2 months ago	
<div></div> aol_Lazio_fty	Physical	Other	druid	<div></div>	2 months ago	
<div></div> aol_Lazio_clcbb	Physical	Other	druid	<div></div>	2 months ago	
<div></div> aol_Lazio_UA	Physical	Other	druid	<div></div>	2 months ago	
<div></div> deusto_campaign_GREENGAGE_App_mission_data	Physical	Other	druid	<div></div>	2 months ago	
<div></div> deusto_campaign_ATMOTUBE_data	Physical	Other	druid	<div></div>	2 months ago	

Figure 57: Superset Dataset view.

In summary, our Apache Superset deployment exhibits robust activity and effective utilization across the project, highlighting its essential role in facilitating user-friendly, insightful data analytics. The current statistics—121 charts, 41 datasets, and 14 dashboards (12 published)—demonstrate substantial usage, continuous data-driven inquiry, and a strong collaborative analytics culture. Future recommendations include periodic audits of dashboard relevance, user feedback collection for usability improvements, and ongoing training to maximize Superset's capabilities.

5.11 Data Quality and Structure Dashboard

As the Data Quality Dashboard is a purely frontend application without a dedicated backend or integrated analytics service, advanced usage statistics could not be captured. However, user engagement can be inferred from authentication data recorded by Keycloak, which shows a total of 745 login events (as of June 2025). This could suggest regular access to the application and indicates meaningful interaction with the system.

5.12 UrbanTEP / VISAT

Currently, GISAT is in the planning phase for integrating usage statistics collection into the UrbanTEP / VISAT tool. While the specifics are yet to be finalized, the potential types of data that might be considered for future collection include Basic Usage Metrics, such as average engagement time, traffic source, the number of accesses to the tool, which could provide a general overview of its utilization.

The methodology for data collection is still under consideration and depends on the format of user stories UrbanTEP / VISAT will deliver. GISAT is exploring various options, including the potential future use of a tool like Google Analytics. However, as of now, no data collection is taking place from the client side, and the specifics will be determined in accordance with data privacy and user consent guidelines.

Given that the integration of a statistics collection mechanism is planned, the process for accessing and interpreting these statistics is still in the conceptual stage. Our approach will prioritize ensuring that any future implementation is in line with user privacy and project requirements.

As the integration of usage statistics collection is planned soon, the current focus is on laying the groundwork for how these statistics might eventually be used to evaluate the tool's impact. Potential considerations could include:

- **Understanding User Engagement:** Future piloting data might help in understanding how frequently and in what ways the tool is being used.

- **Informing Tool Development:** Any future collection of usage data will be aimed at gathering insights to guide the continued development and improvement of the VISAT tool.

5.13 DataHub

Datahub is the leading open-source data catalog helping teams discover, understand, and govern their data assets. Within the scope of this project, the Datahub instance currently ingests and catalogs a total of 31 datasets, 13 dashboards, and 138 charts, for an aggregate of 182 entities (see Figure 58). Although every dataset and dashboard include a descriptive summary (100 %), metadata annotations for ownership, tagging, glossary linkage, and domain assignment remain largely unpopulated: only 19.35 % of datasets carry tags, and none of the entities—datasets, dashboards, or charts—have designated owners, glossary terms, or domain assignments.

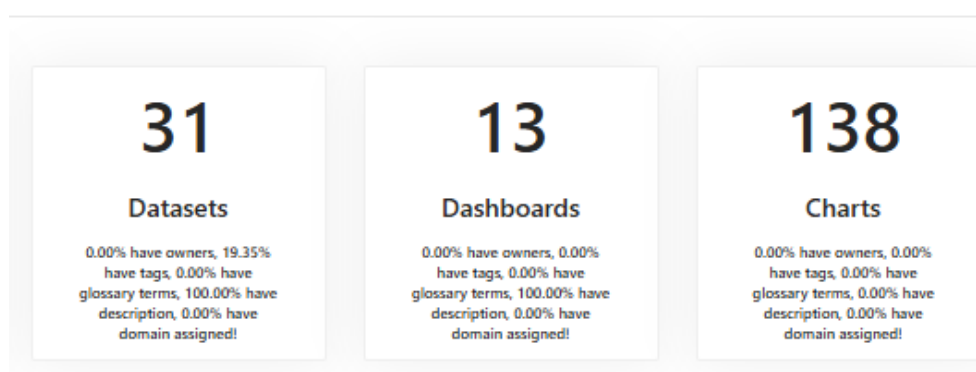


Figure 58: DataHub Analytics panel view.

Entity ingestion is sourced predominantly from two upstream systems. As illustrated in Figure 59 Apache Superset contributes approximately 80 % of the total entities, reflecting its central role as the project's visualization layer, while Apache Druid accounts for the remaining 20 %, representing metric-store metadata ingested into Datahub.

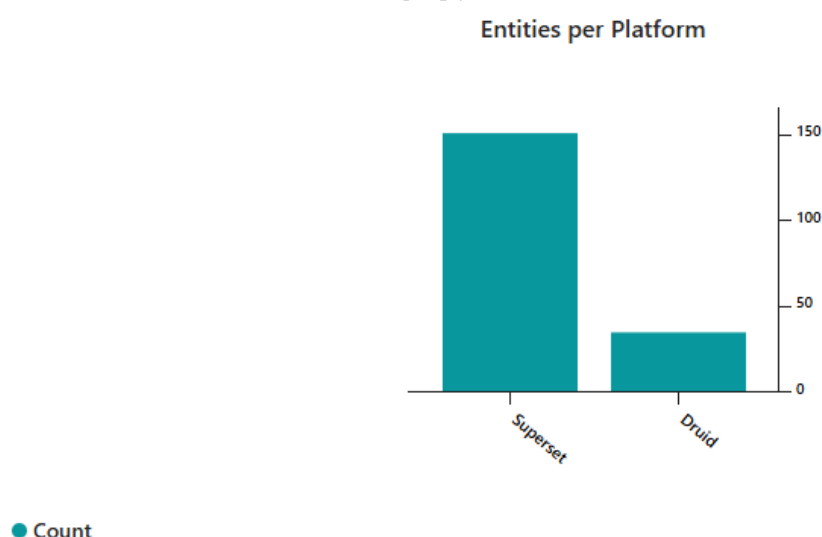


Figure 59: DataHub entities per platform count

These usage metrics reveal clear opportunities to enhance metadata completeness and governance. Assigning data owners will establish accountability and facilitate service-level agreements; applying tags, glossary terms, and domain assignments will improve discoverability and align assets with the project's business taxonomy; and extending descriptions and ownership information to charts will bolster comprehensibility and reuse. By addressing these gaps, the project will fully leverage Datahub's

capabilities to advance data governance maturity and empower users to reliably find and interpret analytics artifacts.

5.14 DigitalTwin

The DIGITWIN itself is not capable of capturing data, nor does the platform itself curate data. On request, curated datasets can be integrated into the twin. However, access is limited to those with a license and login. Once access is granted, through communication with Argaleo, data can be added. However, once added, the analyses can be extracted in Excel charts or graphs.

Most of the analyses stem from the visual combination of different data sets, which require interpretation. The main role of the DIGITWIN then is in Data Exploitation and co-creation. The curation process, as well as the visualized analyses facilitate real life conversation with involved stakeholders – like CO members. Going through the process of exploiting the data for useful insights sets up the possibility for co-creation. The data that must be added will have to be aggregated and analyzed by expert users, but the visualizations will require interpretation by the users. This possibly requires training as well.

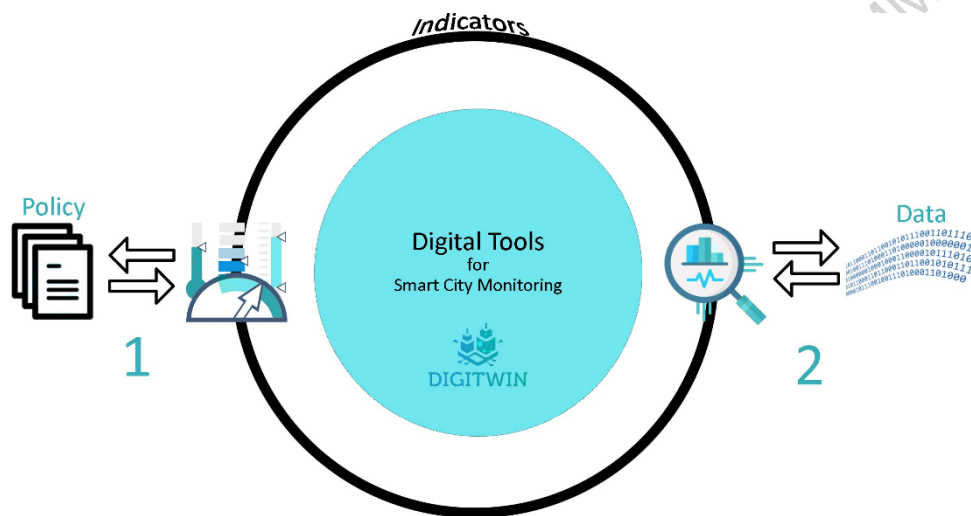


Figure 60: Policy, data and indicator flow chart in DigitalTwin.

The DIGITWIN mostly functions as communication facilitator and data visualization means. However, to measure the impact of the gathered data on the Thematic Co-Exploration of a policy matter, it functions as a middleman between two Citizen Observatory capacities. On the one hand (see Figure 60), the visualizations and their discussions can be related to policy. Combining different data layers and then discussing what they mean with the associated core groups can ensure co-creation on predefined policy matters. On the other hand (indicated by 2), having gathered the data themselves, the citizen observers can explicitly contextualize the visualized data layers, adding more depth and meaning to the gathered data. While not quantifiable in its effect, the DIGITWIN adds qualification to gathered data.

Usage statistics in the DIGITWIN are related to the number of logins provided. While operating with different shells depending on the user, the GREENGAGE shell corresponds to the shell dedicated to the Province of North Brabant. While the main users for this shell for the purpose of GREENGAGE are the involved researchers, the datathon will grant temporary accounts to the participants. This then includes around 30 extra users engaging with GREENGAGE data. The specific GREENGAGE shell for the Digitwin has only been online since June 2025 and the datathon is planned for September 2025. Usage statistics as the time of delivery of this document is therefore preliminary, as the platform is not actively in use yet.

5.15 KeyCloak

Usage Statistics Extraction Statistics are collected using Keycloak's internal event system, which can be configured to store authentication events in its database. Administrators can access this data through the Admin REST API or by querying the underlying event store directly. Furthermore, Keycloak server logs can be configured to provide real-time insights, and metrics can be exported to monitoring systems that support dashboarding and alerting.

In the context of GREENGAGE, the analysis of Keycloak data enables deeper insights into how users are accessing the Citizen Observatories and related tools. For example, it can reveal how many new users registered during a specific campaign, which times and dates had peak login activity, and which authentication methods are preferred or underused. This information can help identify potential security risks, inform decisions around user onboarding, and assess user engagement levels across different system components.

Ultimately, by integrating Keycloak authentication data into the broader analytics architecture of GREENGAGE (e.g., Apache Superset or Apache Druid), project partners can visualize user activity trends, improve platform usability and security, and make informed decisions about feature development and resource allocation.

The following insights are derived from Keycloak's event logs, focusing on authentication errors:

Monthly Trend of Authentication Errors

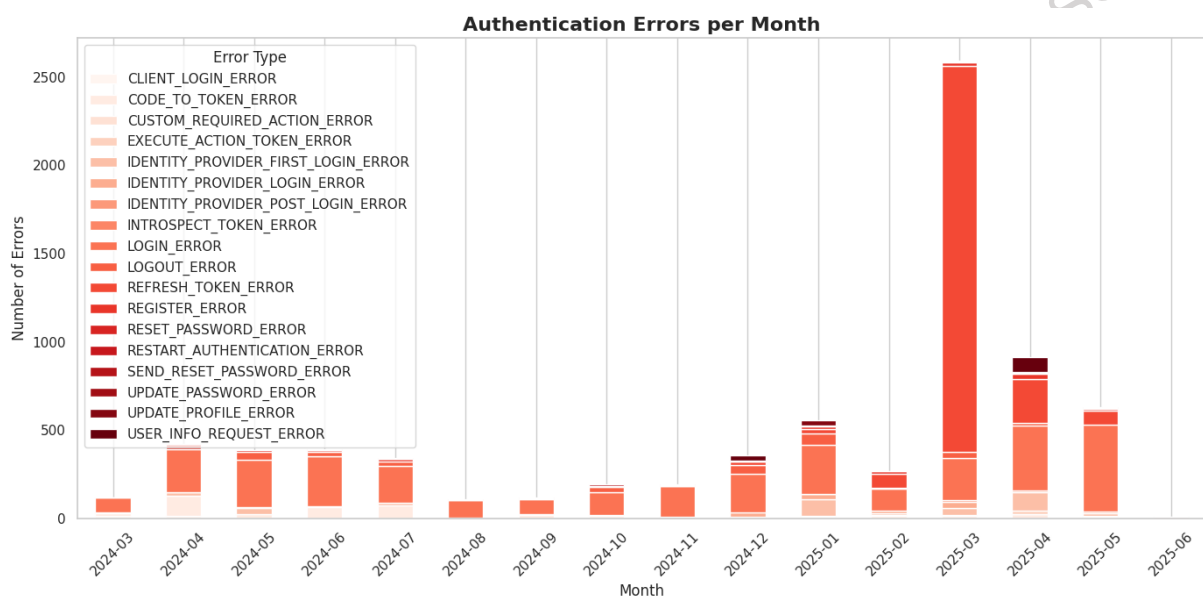


Figure 61: Authentication errors registered by KeyCloak deployment of GREENGAGE.

A time-based visualization highlights the volume and types of authentication errors occurring each month. This helps identify critical periods of system misconfiguration or unauthorized access attempts.

Top 5 Clients with Most Authentication Errors

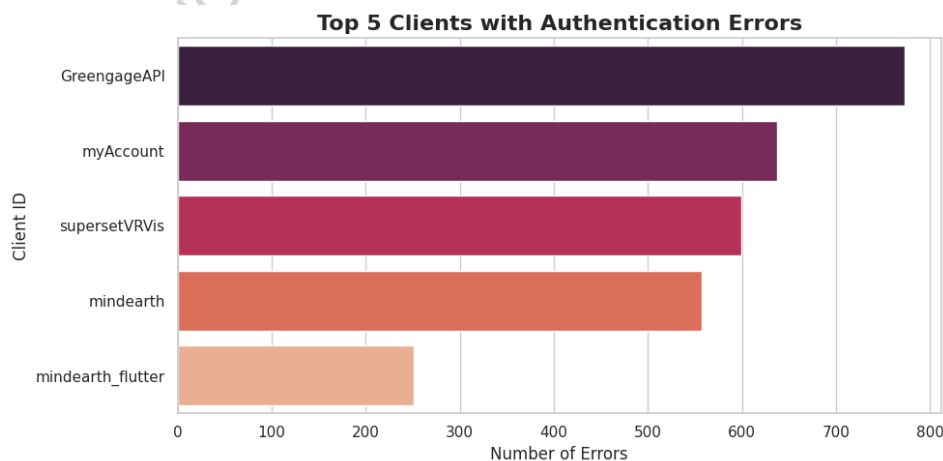


Figure 62: Clients submitting authentication requests to GREENGAGE's KeyCloak-based authentication service.

This chart highlights the client applications with the highest number of failed authentication attempts. A higher error count may indicate misconfigurations or outdated credentials, but it can also reflect more frequent usage—suggesting that these clients are more actively integrated with the system, thereby increasing their exposure to potential authentication issues.

Distribution of Errors by Type

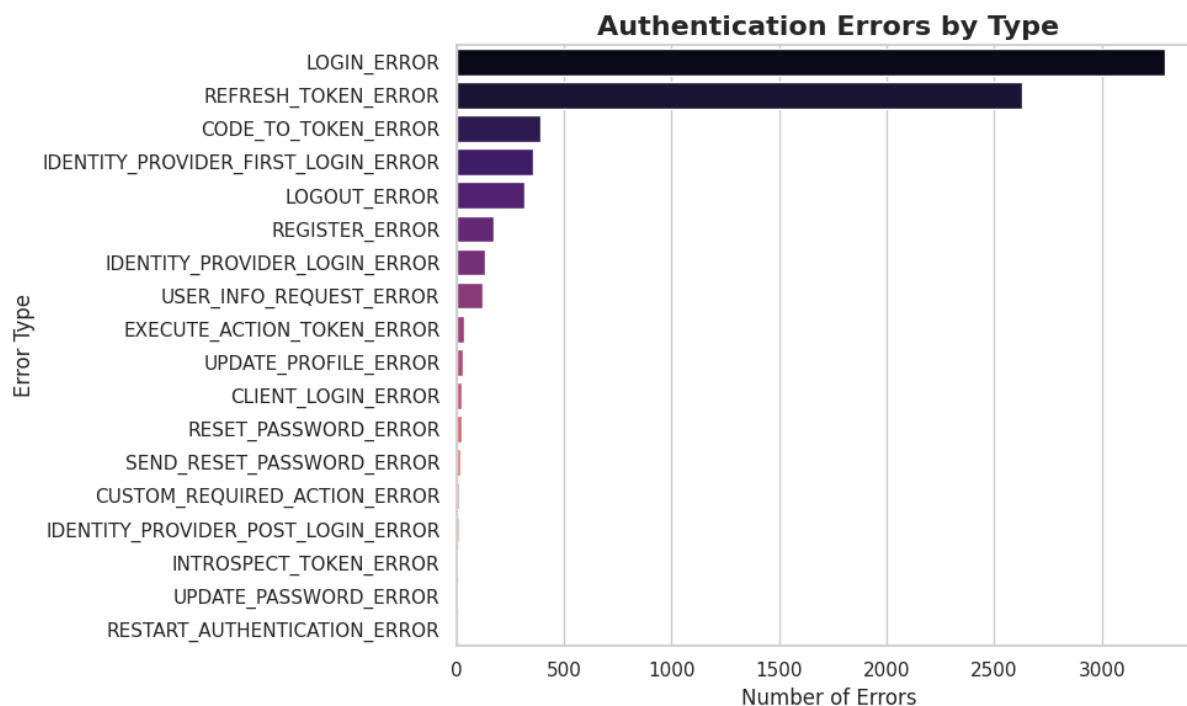


Figure 63: Distribution of errors in GREENGAGE's KeyCloak-based authentication service.

This breakdown categorizes errors into types such as LOGIN_ERROR, REFRESH_TOKEN_ERROR etc., providing insight into whether the issues are user-based, client-based, or protocol-related.

6 Conclusions

This deliverable extends “Deliverable D4.2 – GREEN Engine and manual 2” by updating the descriptions of the tools within GREEN Engine, reporting about usage statistics and providing a new chapter, named “4. Usage manual of GREEN Engine and GREENGAGE’s Academy” where the realization of a full Thematic Co-Exploration through GREENGAGE’s approach and GREEN engine is reported.

The GREENGAGE project, a pioneering endeavour under the Horizon Europe Framework, has made significant strides in engaging citizens in environmental monitoring and policy shaping through the Citizen Observatory Community Journey. This journey represents a structured pathway that empowers citizens and stakeholders to actively engage in environmental observation and decision-making processes. This document, specifically focusing on the manual of the Citizen Observatory Community Journey within the GREENGAGE project, serves as a comprehensive guide outlining the operational aspects of the experiment and monitoring system. It is designed to provide detailed instructions and insights into the implementation and usage of the tools and methodologies developed for the Citizen Observatories. The manual acts as a critical resource for users, facilitating their understanding and engagement in the project, thereby ensuring its effectiveness and success in achieving its objectives.

The Citizen Observatory Community Journey, with its multifaceted phases of Community and Co-production Process Management, Data Capture, Crowdsourcing and Curation, and Thematic Co-Exploration for Insights Generation, encapsulates a dynamic and user-centric approach. The integration of innovative tools and platforms for community, process and holistic data management, such as the Collaborative Environment, GREENGAGE app, MindEarth for GREENGAGE app, and DataHub, underscores our commitment to facilitating a smooth, intuitive, and impactful user experience. These tools ensure easy and efficient flow between areas, enhancing the participatory and data-driven nature of the Citizen Observatories thanks to their ease of use and intuitive design.

Nevertheless, GREENGAGE has faced several challenges in its implementation. These include the consistency of user engagement, data accuracy and reliability, and diverse user needs and Pilot variability. To address these challenges, the project has proposed several complementary solutions, including enhancing interoperability between tools, implementing robust data validation protocols, providing regular feedback and recognition for contributions, and developing a modular approach in our tools and strategies, allowing for flexibility and customization according to local needs and conditions.

In conclusion, the Citizen Observatory Community Journey within the GREENGAGE project represents a well-orchestrated, user-centric approach that seamlessly integrates innovative tools to ensure successful Citizen Observatories and Thematic Co-Explorations. While challenges such as data integration, user engagement, and scalability persist, our proposed solutions aim to address these effectively. The project's approach, grounded in collaboration and innovation, holds the promise of significant contributions to environmental monitoring and sustainable urban development, aligning with the goals of the European Green Deal³¹.

³¹ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en