Distributed Dashboard & Centralised visualisation

Deliverable number: 5.2

Date: 31/05/2024

Version: 1

Name of work package	WP5 Proactive Continuous assessment
Dissemination level	PUB
Due delivery date:	31/05/2024
Actual delivery date:	31/05/2024
Lead beneficiary:	CERTH/ITI
Contributing beneficiaries	-
Authors	CERTH/ITI: Anastasios Pechlivanidis, Avraam Bardos, Evangelos Athanasakis, Theoktisti Marinopoulou, Antonios Lalas, Konstantinos Votis
Peer reviewers	CERTH/HIT: Katerina Touliou, UCD: Francesco Pilla

Executive summary

The deliverable 5.2 of the REALLOCATE project, supported by the European Union's Horizon Europe programme, provides a comprehensive overview of the development and progress of Task 5.2, the AI-powered Distributed Dashboard and Multimodal Data Hub. The main objective of this task is the formulation of the mechanisms and tools to enable the monitoring activities and the definition and development of the related Monitoring System and Open API-tools to enable a continuous monitoring of Key Performance Indicators (KPIs). The deliverable 5.2 which corresponds to milestone at M12, presents the initial version of the distributed dashboard and multimodal data hub with subsequent and final update for M42. Task 5.2 is integral to REALLOCATE, as it develops a comprehensive monitoring and visualisation system that supports the evaluation and impact framework defined in Task 5.1 and informs mid-term assessments in Task 5.3. It also aids in the final impact analysis in Task 5.4 and is essential for Deliverables D5.1 and D5.2, utilising KPIs from Task 2.2 and supporting new mobility services in Task 3.5.

The deliverable explores the architectural framework of the dashboard, highlighting its ability to integrate and analyse data from various pilot sites across cities, calculate each pilot site's KPIs, and continuously monitor them to examine the impact of the cities' interventions. It also emphasises the use of a multimodal data hub where all the data, necessary for the KPIs calculation will be stored. The development of mechanisms to secure user authentication and authorisation is also highlighted. The capabilities and features of the data management platform are described in detail. Additionally, the development of the AI-powered Distributed Dashboard until now, the technologies used, and the way the front-end will display the pilot sites and their relevant KPIs are presented.

Finally, the deliverable outlines the conclusions and the next steps required for the successful completion of this specific task. Additional customisations will be made to the dashboard according to REALLOCATE's needs and requirements. Procedures for data conversion to a uniform format and KPI calculation will be implemented, and adjustments for real-time KPI calculations will be integrated. Additionally, data visualisation tools such as charts, plots, and diagrams derived from the raw data will also be incorporated. The final version will include the live dashboard, with all technical details to be included in the next iteration of the deliverable. This dashboard will allow users to monitor real-time cities KPIs, view statistical information on raw cities data, and compare KPIs between baseline and intervention conditions as well as across different cities.

Table of Contents

List of figures

List of tables

List of acronyms

Acronym	Explanation
AI	Artificial Intelligence
API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
CMS	Content Management System
CSV	Comma-Separated Values
DMP	Data Management Plan
ETL	Extract, Transform and Load
FAIR	Findable, Accessible, Interoperable and Reusable
IoT	Internet of Things
KPI	Key Performance Indicators
MQTT	Message Queuing Telemetry Transport
QoS	Quality of Service
SODA	Socrata Open Data
SQL	Structured Query Language
SSML	Safe and Sustainable Mobility Lab
UI	User Interface
URL	Uniform Resource Locator
WP	Work Package

1 Introduction

This deliverable outlines the progress made in Task 5.2 of Work Package 5 (WP5), focusing on the development of Artificial Intelligence (AI)-based mechanisms and tools for monitoring and assessment. Specifically, Task 5.2 includes the creation of technologies crucial for enabling effective monitoring activities. These technologies include:

- A distributed dashboard and data hub: This component serves as the central point for data aggregation and visualisation, facilitating the real-time tracking of activities and outcomes.
- A comprehensive monitoring system and tools: These are designed to support the continuous monitoring of the Key Performance Indicators (KPIs), ensuring that the project's objectives are being met efficiently and effectively.

The definition of the KPIs, including which indicators the Safe and Sustainable Mobility Labs (SSMLs) will utilise for calculating the relevant KPIs, pertains to Task 5.1. These details will be addressed in Deliverable 5.1, titled "Evaluation & Impact Framework”.

Given that the SSML indicators are scheduled to be finalised by the end of June 2024, and considering that most of the cities have yet to collect data, this deliverable will concentrate mostly on the technologies, architecture, and methodologies applied so far in the development of the multimodal data hub and distributed dashboard. It will also outline the planned approach for calculating the KPIs and how they will be displayed on the dashboard once the cities provide the necessary data.

1.1 Deliverable context

This deliverable reports on the progress of Task 5.2, led by CERTH/ITI with participation from BSC-CNS, SSML cities, and local partners. Task 5.2 is divided into two subtasks: T5.2.1 and T5.2.2, each detailed below.

- **Task 5.2.1** focuses on developing an advanced AI-driven framework to enhance analytics and data handling for REALLOCATE's pilot cities. The aim is to standardise and improve dashboards and data hubs across cities, integrating operational and functional requirements, connectivity, and data models. For cities lacking existing solutions, new dashboards will be introduced, ensuring uniformity across the network. The resulting REALLOCATE Distributed Dashboard and Multimodal Data Hub will feature a comprehensive backend for web services and a unified frontend for visualisation. This system will support real-time predictive analytics and automated data management processes, facilitating informed decision-making for cities.

- **Task 5.2.2** focuses on setting up a sophisticated monitoring system and integrating open application programming interface (API) tools for effective data collection on the KPIs. This effort seeks to build upon and enhance existing infrastructure for optimal data gathering, including the development of tailored software tools for various stakeholder engagements and data types, like environmental and user acceptance. Key initiatives include designing APIs for data integration with the NetZero Cities Platform, improving traffic data collection for non-motorized modes, and implementing strategies for reporting accidents and ensuring accessibility in shared spaces. The task also emphasises establishing secure and efficient data handling protocols, including authentication and pseudonymisation, to maintain user privacy and data integrity.

The Deliverable 5.2 presents the initial version of the distributed dashboard and multimodal data hub at M12 with subsequent and final update for M42. As the project progresses the dashboard will be developed according to the REALLOCATE requirements and SSML needs. The final version of the dashboard will include the real-time visualisation of the KPIs in order to keep a track of the impact of the cities' interventions.

The Deliverable 5.2 corresponds to the milestone at M12, specifically the "REALLOCATE Distributed Dashboard" with the dashboard being live serving as the means of verification.

The REALLOCATE Distributed Dashboard is designed to extract data from each pilot site in real-time, enabling accurate calculation of relevant KPIs. This allows for efficient monitoring and visualisation through the dashboard, facilitating the estimation of interventions' effectiveness and comparison of impacts across different pilot sites.

1.2 Relation with other Tasks and Deliverables

Task 5.2 develops a comprehensive monitoring and visualization system that interfaces with various tasks and deliverables. It builds on the evaluation and impact framework defined in Task 5.1 by leveraging data to measure KPIs and assess the interventions' impacts. The system supports mid-term assessments in Task 5.3, providing essential data and visualisations for identifying corrective actions. Task 5.2 also plays a critical role in the final impact assessment in Task 5.4, facilitating a thorough analysis through comprehensive data visualisation. Additionally, it directly contributes to deliverables D5.1 and D5.2 by defining and materialising project KPIs and detailing the distributed dashboards and centralised visualisation tools. By utilising KPIs from Task 2.2 and supporting the implementation of new mobility services in Task 3.5, Task 5.2 ensures effective tracking and strategic recommendations for SSMLs. Furthermore, it enhances WP4 by facilitating peer learning and capacity building through shared visualisations and best practices, and supports WP6's communication and dissemination efforts by providing a centralised platform for showcasing project outcomes and engaging stakeholder, ensuring a cohesive and impactful implementation of the project's objectives.

1.3 Deliverable structure

This document is structured as follows to provide a comprehensive overview of the task's scope and methodologies:

- **Section 2** explores the range of technologies documented in current literature that are relevant to monitoring activities and data hubs.
- **Section 3** delves into the design principles and architectural considerations that underpin the dashboard, laying the groundwork for its functionality and user interface.
- **Section 4** provides an in-depth look at the technical requirements for installing and making CKAN, the chosen data repository, accessible to the public. This section details both the software and hardware dependencies necessary for deployment.
- **Section 5** describes the specific dependencies required for the dashboard installation, detailing the software and configuration settings essential for its setup and optimal performance.
- **Section 6** anticipates the project's trajectory, highlighting forthcoming tasks and objectives. It concludes with a reflection on the document's findings, underscoring key insights and potential implications for future development.

2 State-of-the-art technologies

In the previous years, intensive efforts have been made, to bring the concept of smart cities from vision to reality. In this context, the tools and technologies for retrieving, storing and sharing data originating from smart city sensors, as well as visualising these data, play a crucial role. To this end, a state-of-the-art analysis is presented in this section. The Scopus[1] and IEEExplore[2] databases were used along with the following search query:

"(urban OR smart AND city) AND data AND management AND transportation AND analytics AND environmental AND analytics"

According to a literature review paper related to big data in smart cities [1], vehicular traffic data is an important part of advanced cities, while considering climate data can furthermore enhance its value. The authors emphasise the use of tools like MATLAB[3], Python[4] MapReduce and Hadoop[5], in most applications. In closely related work, Pereira et al. [2] demonstrate a decision support dashboard for the city of Lisbon, Portugal. Their solution utilises data from three different sources and, after appropriate pre-processing, stores them in a PostgreSQL database. The APIs exposed based on this database are used to render a visualisation dashboard.

In another recent review paper, which focuses more on the algorithms used to preprocess or to utilise big data in smart city applications [3], it is emphasised that the forecasting of social phenomena based on city data is the most desirable result. Nevertheless, it still remains as a major challenge, as the whole pipeline of the assembly, processing, and analysis is important. They also state that out of the 20 selected papers, only six include a data pre-processing and analysis framework. In [4] Joyce et al. studied the data protection and sharing strategies required in smart cities and proposed a universal framework to achieve it. The framework emphasises the following aspects: 1) Strategic Planning and Stakeholder Alignment, 2) Data Inventory, Standards, Compatibility and Quality, 3) Data Governance, Privacy and Security Measures, 4) Compliance with Data Protection Legislations, 5) Data Sharing Agreements and Ethical Considerations, 6) Sharing Infrastructure and Technology Solutions, 7) Monitoring, Evaluation, and Communication. The authors also highlight as potential future work, the exploration of data standards with the help of AI and the

[1] https://www.scopus.com/search/form.uri?display=basic&zone=header&origin=recordpage#basic
[2] https://ieeexplore.ieee.org/Xplore/home.jsp
[3] https://www.mathworks.com/products/matlab.html
[4] https://www.python.org/
[5] https://hadoop.apache.org/

investigation of common data sharing, in specific sectors such as transportation, to provide guidance for each sector's unique requirements.

Another comparative study [5] focuses solely on the exploitation of open data in the context of smart cities. They claim that Comprehensive Knowledge Archive Network (CKAN)[6], is the most widely used and acceptable tool, for managing open data. Nevertheless, a major challenge remains in automating the Extract, Transform and Load (ETL) processes. The authors compare different ETL solutions developed on top of CKAN, and emphasise that none of them enable non-IT personnel in the open data lifecycle, since their exploitation requires code or Structured Query Language (SQL) queries. This issue is highlighted as a future work direction. Liu et al. [6] propose a framework called CITIESData, for the management of the heterogeneous smart city data, by defining three data categories, sensitive, quasi-sensitive, and open/public levels. The overall architecture of CITIESdata is illustrated in Figure 1. In this architecture the final storage layer is an open data system, like CKAN, Zenodo[7] or DataCite[8]. In this layer, only open data, or public metadata of the quasi-sensitive data, can be stored. Nevertheless, sensitive or quasi-sensitive data, can be transformed into open data by undergoing anonymisation processes through their developed big data processing pipeline called bigETL.

In another notable study [7], the authors present a visualisation tool that allows the fusion and extraction of heterogeneous smart city data by analysts. Though the tool is well presented and its capabilities seem promising the authors discuss little about the technologies used to store and manage the data in the back end. On the other hand, Suciu et al. [8] present an end-to-end architecture for managing and exploiting heterogenous data in smart cities, by exploiting the Message Queuing Telemetry Transport (MQTT) protocol. Even though the description of the high-level architecture is provided, it is not backed by instructions on how it could be implemented, or the report of any evaluation metrics. Similarly, [9] present a methodology for streaming and analysing Internet of Things (IoT) data in real time and back it up by an anomaly detection application for environmental data. They claim their methodology is domain independent and can be applied to different types of data. In another noble solution [10], the authors emphasise issues like quality of service (QoS) and privacy in smart city data related to transportation. They have used cloud technologies like Apache Spark to address them. On top of this solution, Fiore et al. [11] have built a city administration dashboard with emphasis on transportation.

[6] https://ckan.org/
[7] https://zenodo.org/
[8] https://datacite.org/

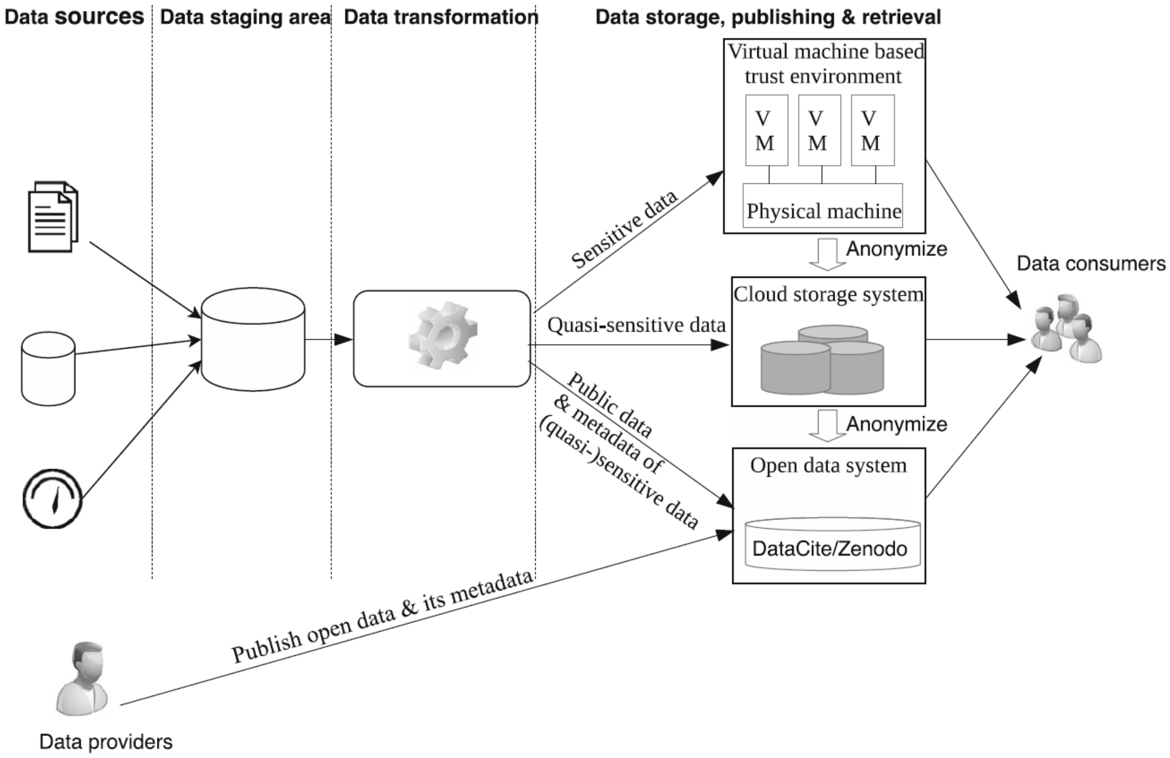

Figure 1. Architecture of CITIESdata [6].

Based on the insights from this state-of-the-art analysis, it is concluded that a data management tool like CKAN would be useful for the needs of REALLOCATE. Since most of the expected data are open and localised in specific sections of the city, big data platforms like Hadoop would add more complexity with little value to the solution. Finally, three pillars are recognised as providing significant additional value to the data lifecycle of REALLOCATE:

1. The enhancement of the management platform with an ETL process to alleviate potential interoperability and data protection concerns.
2. The exposure of APIs that will make the stored data available.
3. The development of visualisation dashboard where policy makers can have a high-level overview of the data.

3 Architecture

This section introduces the preliminary architecture of the REALLOCATE Distributed Dashboard, which is currently composed of three principal elements: a) the pilot sites, b) the data management platform, and c) the dashboard interface. As the project progresses, changes or additions may arise in this architecture.

Each pilot site, represented by a city, is responsible for supplying historical data essential for the calculation of the KPIs through the CKAN platform. This platform and the capabilities it enables will be discussed in section 4 in more detail. Authorised city representatives will be granted access by CERTH to CKAN for the purpose of dataset uploads. The implementation of Keycloak is mandatory for managing and securing user authentication and authorisation, ensuring that only authorised city representatives can access and manage datasets on the CKAN platform.

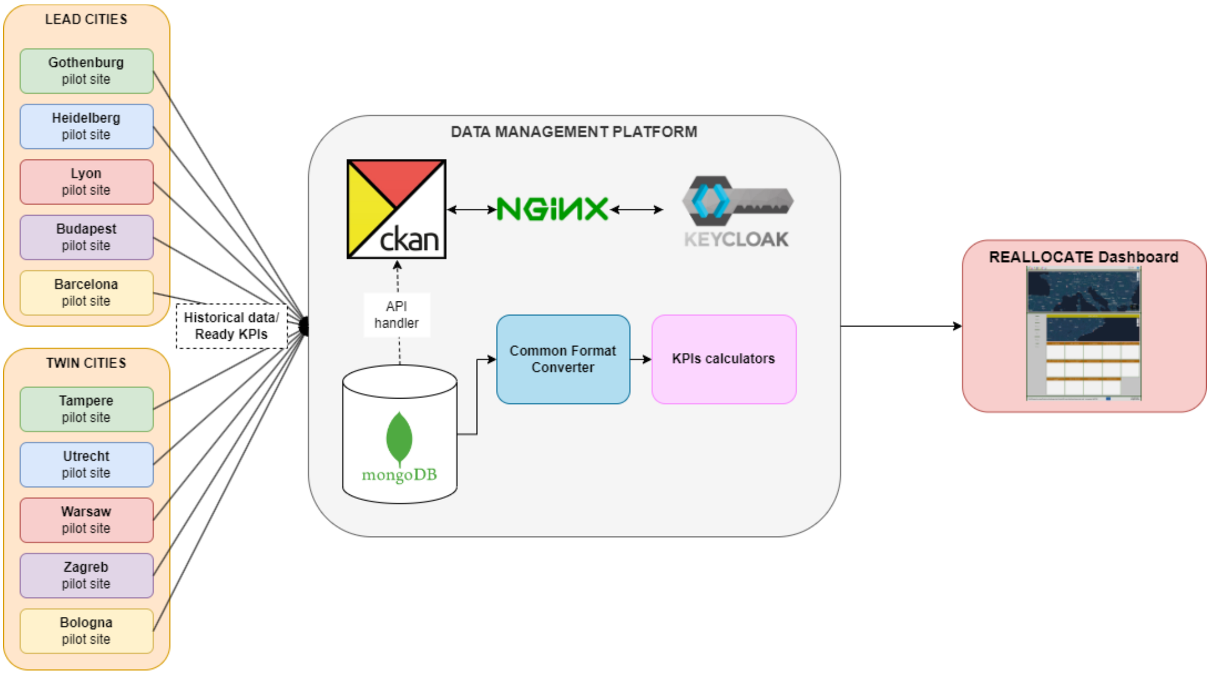

Figure 2. Architecture of REALLOCATE's AI-powered distributed dashboard.

The cities may provide either pre-calculated KPIs or the raw data required for computing these KPIs. This data will be stored in a MongoDB database that is integrated with CKAN. While a common format such as CSV is preferred for data uniformity, CERTH will undertake the conversion of data from incompatible formats. For the KPIs that are not pre-calculated, CERTH will perform the necessary calculations using methodologies best suited for each specific KPI. Common data types are necessary for the common KPIs across cities to enable better comparisons and the use of standardised calculation methods.

The resulting KPIs will then be displayed on the dashboard. The dashboard will provide a map-based interface that allows users to view the location of pilot sites, their active status, as well as their respective KPIs, enabling comparisons across cities and the impact of the applied interventions. Figure 2 illustrates the architecture described. In the final version the data will be sent in real-time using MQTT messaging protocol. Pilot sites will be able to connect to an MQTT broker and publish the available data to the corresponding MQTT topics. The aforementioned procedure will be analysed in the next version of the document.

Finally, the dashboard will provide smart visualisation tools, including diagrams and statistics, derived from the data collected at the pilot sites. All this information will also be included in the dashboard and developed by BSC-CNS.

4 Multimodal Data Hub

This section outlines the Multimodal Data Hub, explaining the necessity for its development. It details the data and platform requirements and compares various data management platforms, ultimately concluding that CKAN is the optimal choice. It thoroughly presents all features of CKAN, including user login and registration, organization creation, dataset storage, and visualisation.

4.1 Role overview

The Data Collection Platform utilised for REALLOCATE plays a pivotal role since it is necessary for storing all the data collected for the KPIs calculation. It should be able to handle and store a variety of data from all pilot cities and not limit their available data to only a particular format. This variety of data types is often referred to as multi-modal data [12], denoting that various types of data (tabular, audio, image, video etc.) are being obtained through multiple modalities, and not from a single source. The place that these data are stored is considered a multi-modal data hub, and presents both opportunities and challenges for organizations seeking to harness its potential use cases and applications.

Unlike traditional data management systems that are limited to specific data types and formats (require a specific schema for the data to be saved as), a multi-modal data hub is designed to accommodate, store, manage and analyse diverse types of data. For instance, since a multi-modal data hub does not require any particular schema for the data, both structured and unstructured data regardless their type (CSV, JSON, PDF etc.), can be handled [13]. This makes the multi-modal data hub specifically advantageous in the case of REALLOCATE as it involves multiple data sources across all the cities that collect various types of data.

4.2 Requirements

This section presents the requirements that the multi-modal data hub should adhere to. Specifically, it is divided into two parts: the data requirements, which include the principles that the data hub should follow regarding its stored data, and platform requirements which specify the general requirements and functionalities that the data hub should provide.

4.2.1 Data Requirements

Attempting to enhance the ability of both human operators and automated systems to efficiently discover, access, integrate and reuse data, the introduction of the Findable, Accessible, Interoperable and Reusable (FAIR) principles [14], has arisen as a pivotal advancement. The REALLOCATE's Data Management Plan (DMP) is also developed in a way to adhere and follow the FAIR principles, and a detailed description of the aforementioned principles is provided as follows:

- **Findable:** Each dataset should be assigned with a unique identifier (to ensure long-term discoverability), rich metadata (information about the content, structure and relevant attributes), rich keywords (to enhance searchability within the community) and indexed (searchable through repositories or search engines).
- **Accessible:** Each dataset should be easily accessible and retrievable by their unique identifier through a universally implemented protocol. This ensures that retrieval of data can be facilitated, with provision for authentication and authorization procedures to safeguard data integrity and privacy when necessary.
- **Interoperable:** The representation of each dataset, and its accompanying metadata, should be performed in a standardized, formal and broadly applicable/adaptable language. Furthermore, the establishment of linkages between these datasets via metadata is imperative for comprehensive data integration. It is essential that the vocabulary employed within metadata also adheres to the FAIR principles, ensuring robustness and accessibility across diverse contexts.
- **Reusable:** Each dataset should be accompanied by a comprehensive documentation, easily accessible data usage licenses, and thorough provenance information. This ensures transparency and facilitates reproducibility, vital elements in fostering robust data practices.

4.2.2 Platform Requirements

The subsequent requirements specify the fundamental aspects that the multi-modal data hub should encompass as a platform, aimed at fostering and creating a resilient, robust and user-friendly environment for managing diverse data effectively. Each requirement assumes a vital role in shaping the overall performance of the platform, and by adhering to these requirements the multi-modal data hub will be able to serve as a valuable resource facilitating collaboration and innovation across various domains. The main platform requirements are:

- **Reliability:** The Platform must exhibit robust performance across diverse conditions to ensure operational integrity and data accessibility. This entails the minimization of downtime to maintain continuous service provision, thereby, ensuring reliability through sustained availability and uninterrupted data access.
- **Maintainability:** Comprehensive documentation detailing the platform's architecture and the available features it offers should be provided. This procedure helps streamline future maintenance efforts and enhance the efficiency of troubleshooting procedures.
- **Extensibility:** The platform should demonstrate a robust capacity for accommodating the seamless integration of supplementary features and functionalities through well-defined extension mechanisms. Equally, developers should have access to developmental resources, such as thorough documentation, facilitating the process of customizing the platform to meet evolving needs.
- **Robustness:** The platform should incorporate robust user input validation and error handling measures to prevent and mitigate security breaches and vulnerabilities, while minimizing potential data integrity issues.
- **User-friendly:** The platform's interface should embody a high degree of user-friendliness, characterized by responsiveness and intuitive design. Such attributes are indispensable for accommodating users across a varying level of technical expertise, ensuring seamless navigation and accessibility.
- **Security:** The platform must adhere to industry-standard protocols to safeguard against potential threats. This entails the implementation of both access controls and authentication mechanisms.
- **Accessibility:** The platform should be accessible via a dedicated internet domain name to ensure widespread reach and availability. Moreover, the platform should exhibit responsiveness and compatibility across a wide spectrum of devices, thereby accommodating diverse user preferences and technological environments.

4.3 Data repository

A critical element of the multi-modal data hub infrastructure is the data repository component, which functions as the fundamental framework for the storage, administration, and dissemination of multi-modal datasets across all involved parties. To fulfil this essential role, the CKAN[6], distinguished for its adaptability and resilient functionalities, has been selected as the preferred open-source solution for the data management platform.

Due to its proven effectiveness in facilitating transparency, fostering data sharing mechanisms and collaboration capabilities, CKAN has garnered significant traction across various sectors, spanning commercial enterprises and governmental entities alike. Prominent instances of CKAN deployment, exemplifying the wide-ranging applications within global landscape of data management and dissemination, encompass the United States' (U.S.) federal data portal[9], managed by the U.S. government, the Canada data portal[10] maintained by the Canadian administration, and the European data portal[11].

It is important to mention that the data repository will function as a centralised data hub for all cities, facilitating easy access and integration of multi-modal data while ensuring that all cities are provided with the means to securely share their data. Some cities, such as Tampere[12], Lyon[13], and Barcelona[14], may already have their own data repository. Regardless of whether a city has an existing repository, all relevant datasets should also be included in the centralised data hub, with the only difference being that, cities with their own data repository will have an additional option when sharing data on the centralised data hub.

Specifically, cities without an existing data repository will directly upload their data files to this centralised repository by following the process outlined in Section 4.4.3. On the other hand, cities that already have an existing data repository are also required to include their datasets in this centralised repository, but rather than duplicating the data by uploading the file itself, these cities can opt to provide a direct link to the existing dataset (Figure 9). This approach promotes consistency across all involved parties and reduces redundancy while ensuring that the platform provides comprehensive data accessibility. Finally, a user manual will be also provided to guide cities through all functionalities of the data hub, ensuring easy and efficient use of the platform.

---

[9] https://data.gov
[10] https://www.canada.ca/en.html
[11] https://data.europa.eu/en
[12] https://data.tampere.fi/fi/
[13] https://data.grandlyon.com/portail/fr/accueil
[14] https://opendata-ajuntament.barcelona.cat/en

4.3.1 Data Management Platforms

Several alternative data management platforms also merit consideration. Among these, platforms such as DKAN[15], Socrata[16] and Junar[17] offer functionalities similar to those of CKAN and have significant adoption rates in data management and dissemination applications. In the following sections, an exhaustive exposition of each one of these platforms is presented, showcasing their respective utilities and features. This analysis is focused on provided a detailed assessment of the CKAN's essential features that make it the preferred choice for the data management platform.

4.3.1.1 CKAN

CKAN[6] stands as a robust and adaptable open-source data management system, empowering organizations with functionalities and tools to effectively store, manage and seamlessly share data among different partners. Given its open-source nature, and a very active community dedicated to its development and advancement, CKAN is considered a secure solution, favored by numerous regional government organizations throughout the globe. This widespread adoption underscores not only its resilience to vulnerabilities, but also encourages the creation of an engaged community, committed to ensuring CKAN remains up-to-date and vigilant against emerging security challenges, thus safeguarding data for future threats.

Similar to REALLOCATE, several projects harness the capabilities of CKAN as their data management platform, facilitating not only data storage and management, but also enabling useful data visualisations to better understand data structure and content. Specifically, Energi Data Service[18], serves as a free and open data portal that utilizes CKAN to store data focusing on Danish energy-related information, including production data and CO2 emissions. Additionally, an intriguing case is presented by The Humanitarian Data Exchange (HDX)[19] which leverages CKAN as a feature that allows organizations to share their work and give visibility to their sensitive data through visualisations, all while maintaining strict data privacy measures and without the need to share the actual data.

CKAN offers an extensive documentation and multiple ways of installation to accommodate the requirements of diverse projects. Particularly, installation of CKAN can be performed through multiple ways, including package installation (suitable for single-website server

[15] https://getdkan.org/
[16] http://open-source.socrata.com/
[17] https://junar.com/
[18] https://www.energidataservice.dk/
[19] https://data.humdata.org/

deployment for quick and easy implementation), source installation (optimal for development purposes) and Docker installation (requires less setup effort than a source install and more flexibility than a package install). In addition to these installation options, some of the core functionalities of CKAN are:

- CKAN's Action API[20]: a powerful API that exposes all of CKAN's features beyond its conventional web interface. Empowering applications with seamless integration, the API facilitates the retrieval of JSON-formatted list of datasets and groups, as well as JSON representations of datasets and resources. Furthermore, actions regarding data management operations such as creating, updating and deleting datasets can be performed. Beyond these actions, it also allows advance search capabilities and grants access to activity streams that showcase recently performed activities.
- CKAN's DataStore[21]: an extension that serves as a flexible database solution. By integrating resources into the DataStore, users can seamlessly filter and update data without the need to resort in time consuming solutions such as downloading and re-uploading data. Moreover, users gain access to the searching API and authorization system, and also get automatic data previews by leveraging extensions such as Data Explorer[22], enhancing the efficiency and usability of the system.
- Extensibility: With a repository composed with more than 200 community extensions, CKAN offers the flexibility to users to customize and tailor any data portal to meet the most of requirements. Furthermore, by taking advantage of the extending guide[23], users have the ability to extend existing features, or even develop custom ones based on their needs.
- Multi-modality: CKAN effectively structures data within discrete "datasets" entities. Each dataset comprises of two components, metadata and resources. Metadata include data details ranging from the title and unique identifier, to detailed descriptions, available formats and licensing information. Meanwhile, the resources component serves as the repository the actual data, regardless their format (CSV, XML, PDF, image etc.).

[20] https://ckan.org/features/api
[21] https://ckan.org/features/datastore
[22] https://docs.ckan.org/en/2.9/maintaining/data-viewer.html#data-explorer
[23] https://docs.ckan.org/en/2.9/extensions/index.html

4.3.1.2 DKAN

Inspired by CKAN, which was introduced and analysed in section 4.3.1, DKAN[24] emerges as a distinguished open-source and open-data platform built on top of the Drupal Content Management System[25] (CMS). DKAN's modules and subsystems are structured around four pivotal pillars of data functions: Management, Aggregation, Discoverability and Usability. A description of these function is provided as follows:

- Management: as outlined by DKAN, data management entails strategies for data governance. This includes empowering and equipping data publishers with the tools to streamline data dissemination, and thereby, aims to empower data consumers to effectively discover and utilize data pertinent to their objective. However, while making data public can be straightforward, ensuring data openness (making data accessible, discoverable, machine-readable, interconnected with contextual resources and appropriately licensed) requires meticulous effort, which can be drastically reduced with the infrastructure provided by DKAN.
- Aggregation: a key feature within the DKAN platform, is the ability to host aggregated or federated datasets alongside originally sourced data. This process, referred to as harvesting, constitutes an integral component of the Harvest[26] module offered by DKAN.
- Discoverability: since the usefulness of data is directly impacted from how easy it is to access and comprehend them, DKAN houses dedicated modules designed to enhance data discoverability. Notably, DKAN Metastore empowers publishers to enrich datasets with contextual metadata, and in turns, make data searchable through search mechanisms.
- Usability: typically, data stored within files lack searchability due to inherent limitations. Leveraging the DKAN Datastore resolves this challenge by adeptly parsing and storing data in a format conductive to exploration, and allows direct access to the stored data.

To enable the aforementioned functionalities, DKAN employs four primary components: Metastore, Datastore, Harvest and REST API. The Metastore is used to create, retrieve, update and delete information describing the data, also known as metadata. The Datastore provides services to parse data (predominantly in CSV format) and store them into database tables, allowing users to query data via a public API. The Harvest module enables the

[24] https://dkan.readthedocs.io/en/latest/index.html
[25] https://www.drupal.org/
[26] https://dkan.readthedocs.io/en/latest/components/dkan_harvest.html

seamless communication with other external data portals, allowing users to import items from these portals through either public feeds or APIs.

4.3.1.3 Socrata

The Socrata[16] open data value framework, is a cloud-based software enterprise committed to broadening the accessibility of public data across diverse organizations globally, including prominent urban centers such as New York and San Francisco. At its core, Socrata aims to enhance transparency, streamline citizen access to information and integrate data into decision-making processes cost-effectively. This service helps liberate data from corporate barriers, and initiates the conversion of raw data into actionable insights, making them universally accessible, visualisable and readily shareable.

Socrata, as an open data value framework, encompasses four fundamental pillars[27]. Firstly, it prioritizes citizen experience, ensuring transparent and high-quality user interactions that enhances the quality of life through open data initiatives. Secondly, it promotes data-driven decision making, providing improved governance through measurable performance tracking. Thirdly, it emphasises operational efficiency, enabling organizations to optimise processes by leveraging cloud-based infrastructures. Lastly, it underscores economic impact, as open data applications have demonstrated tangible contributions to local economies as a result of open data applications.

A key component of this platform, is the Socrata Open Data API (SODA API) [15], a set of RESTful APIs designed to empower developers with easy access and manipulation capabilities for datasets housed within the Socrata ecosystem. Specifically, the SODA API provides endpoints for querying, filtering and data manipulation, equipping developers with the essential tools required to built open data applications.

4.3.1.4 Junar

Junar[17] is a cloud-based data platform that empowers organizations to publish and share data resources effectively. Its main objective is to convert existing datasets into actionable insights through dynamic tables, visualisations and even API's, simplifying data accessibility for developers. At its core, Junar supports four functionalities categorized under the topics of: datasets, data views, visualisations and dashboards. Specifically:

- Dataset is a resource that can be obtained from a document hosted either on you computer, or through a web service, accommodating various file formats.

[27]https://static.carahsoft.com/concrete/files/5314/2748/7671/Socrata_Whitepaper_The_Open_Data_Value_Framework_1.pdf

- Data view emerges as a distinct entity created from a dataset, representing either a subset of the dataset or individual records of it. While data views operate independently from datasets, modifications made to the dataset trigger automatic updates in corresponding data views.
- Visualisation is a graphical representation of selected data extracted from a data view. These representations can be either tables, charts or maps.
- Dashboard is a collection of data views and visualisations, providing a complete overview of data insights.

Access to Junar is granted though a user authentication process, ensuring only individuals with valid credentials gain entry. Each user is allocated a distinct role among three tiers: editor, publisher and administrator. An editor has the ability to create, edit and send datasets, data views and visualisations for revision. Publishers have the same privileges to editors, with the added authority to publish, unpublish and delete resources. Administrator is the highest tier role with similar privileges to publishers, while also having the ability to alter and make changes over the platform's configuration.

4.3.2 Comparison of Data Management Platforms

The focus of this section is to outline and provide a comprehensive comparison with the strengths and weaknesses of the aforementioned data management platforms (CKAN, DKAN, Socrata and Junar). Various critical aspects are used to determine which data management platform best covers the requirements outlined in Section 4.2, and thus, is superior and more suitable for the needs of REALLOCATE. Table 1 provides a clear overview of the supported functionalities of each data platform, aiding this way in the selection of the most qualified platform.

Table 1. Comparison of Data Management Platforms.

Feature/Aspect	CKAN	DKAN	Socrata	Junar
Open-source	Yes	Yes	No	No
Community Support	Large/active community	Small community	Limited	Limited
Customizability	Highly customizable	Customizable	Limited customization	Limited customization

Feature/Aspect	CKAN	DKAN	Socrata	Junar
Multi-modal Data Support	Yes	Yes	Yes (limited)	Limited
Metadata Support	Strong	Basic	Strong	Basic
Data Visualisation	Strong	Basic	Strong	Basic
Search Functionality	Strong	Good	Strong	Basic
Extensibility	Strong	Good	Limited	Limited
Data Governance	Strong	Good	Good	Basic
Security	Strong	Good	Strong (Enterprise)	Basic
Reliability	Strong	Good	Strong	Basic
User Interface	User-friendly	User-friendly	User-friendly and intuitive	Basic interface
Installation/ Deployment	Flexible	Flexible	Primarily cloud-based	Primarily cloud-based
Documentation	Strong	Good	Good	Limited
Language Support	Strong	Good	Good	Limited
API Access	Strong	Good	Strong	Good
Cost	Free	Free	Subscription-based	Subscription-based

Based on Table 1, CKAN clearly outperforms DKAN and Junar by offering all of their functionalities along with greater flexibility and additional features. Compared to Socrata, which provides strong data visualisation methods and is highly reliable, CKAN, as an open-source solution, offers more flexibility in both customization and installation processes while being supported by a large and active community. These factors make CKAN the superior choice for meeting the needs of the REALLOCATE data management platform.

4.4 CKAN as a portal for REALLOCATE

This section aims at illustrating the operational capabilities of the first version of CKAN, employed as a data portal for REALLOCATE, via its accessible User Interface (UI). CKAN inherently furnishes a comprehensive and ready to use web interface, affording users the ability to seamlessly engage with its suite of features and tools. This default interface is illustrated in Figure 3.

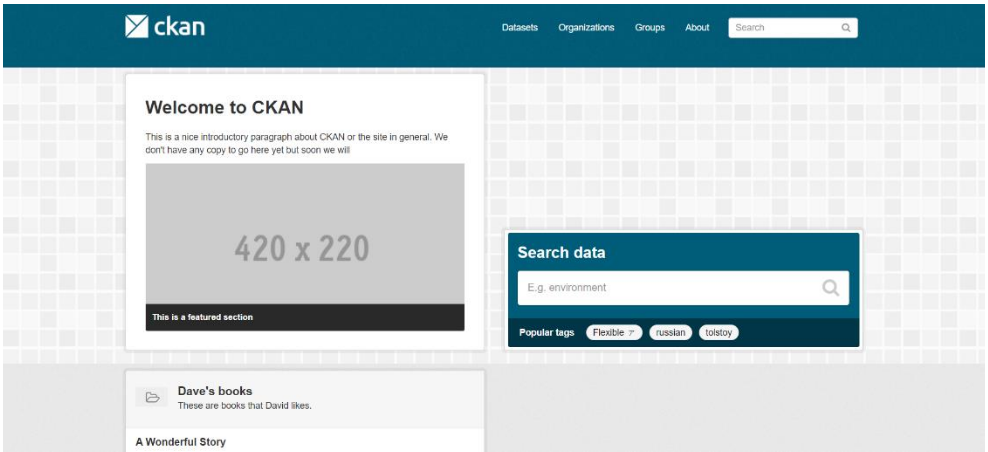

Figure 3.The default CKAN web page.

To further facilitate the needs of REALLOCATE, additional enhancements were made to the UI of CKAN, aimed at achieving a more user-friendly and familiar to all involved partners web interface. The tailored version of the CKAN which serves as REALLOCATE's designated data repository and provides functionalities such as seamless user login, data upload and data retrieval through search mechanism can be seen in Figure 4.

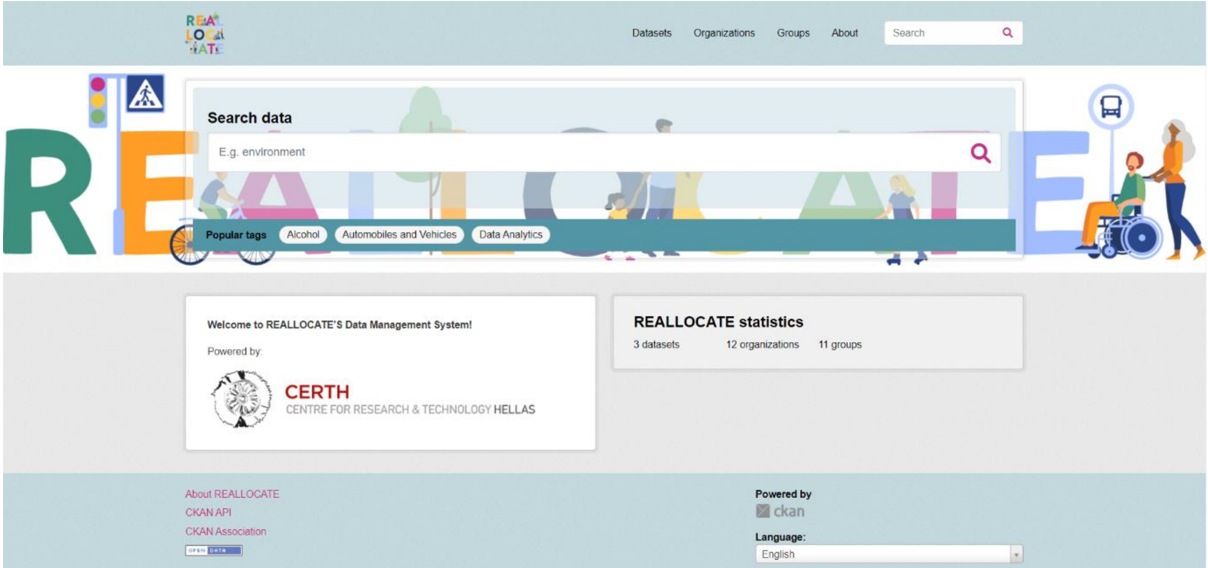

Figure 4. REALLOCATE's data repository.

4.4.1 Login and Registration page

Upon accessing the data repository of REALLOCATE, users will utilise their individual credentials for authentication and secure login. Each partner will receive distinct credentials, eliminating the necessity of registration, which has been removed from the UI, boosting this way security even more, since no external individuals will be able to interact with the repository. The login page interface is illustrated in Figure 5, where users are prompted to enter their username or email address, accompanied by their password in order to gain access to the repository. Should a user encounter password-related issues or difficulties logging in, the "Forgot your password" feature provides account recovery functionality, where by supplying the associated account email address or username, users will receive an email containing detailed instructions for account retrieval.

4.4.2 Organizations in CKAN

Organizations serve as the primary mechanism for managing access to datasets within the CKAN platform. Each dataset is exclusively associated with a single organization, which in turns, determines authorized to view, create and modify its content. The association between data and organizations occurs either when someone uploads a dataset directly through the organization's page or by choosing to assign the dataset into a specific organization when uploading from the dataset's page.

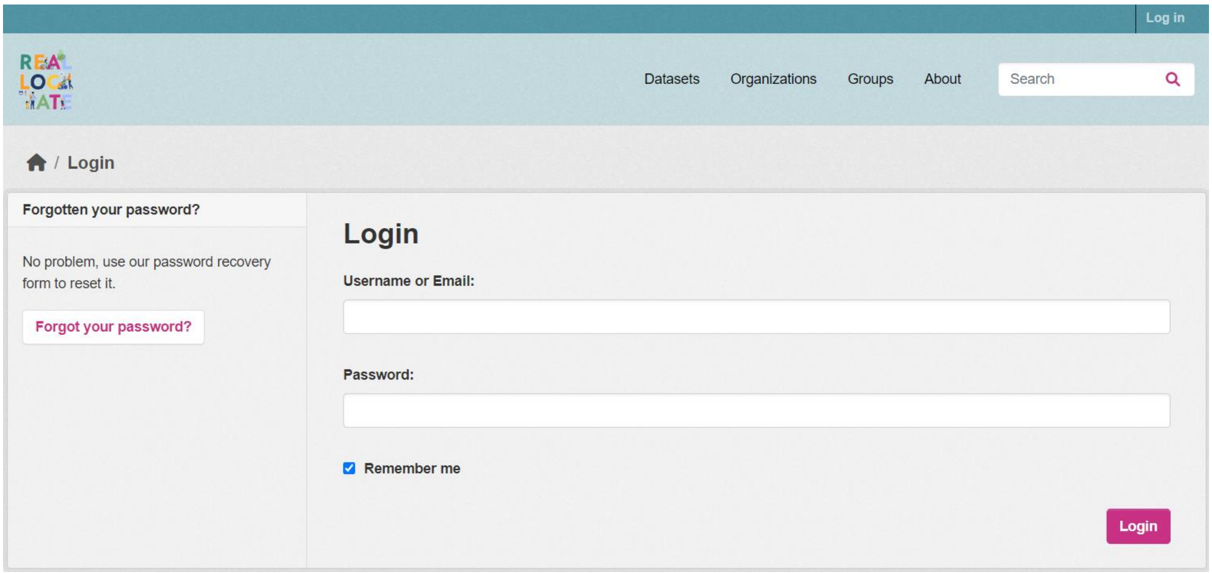

Figure 5. Login page in REALLOCATE's data repository.

Furthermore, during dataset creation, a categorisation between public and private should be made. Public, is a dataset the can be visible by any user, while private datasets have a restricted visibility to solely authenticated users who also hold a membership within the respective organization that the dataset is assigned to. It is important to mention that private datasets are also excluded from the general dataset searches, but remain visible in dataset searches within the organization. The organization page of the REALLOCATE data repository is depicted in the Figure 6. It is important to note that these organizations are subject to potential modifications and enhancements in subsequent iterations.

Within organizations, individuals get assigned into one of the three distinct roles: member, editor or admin. Upon the creation of a new organization by a user, the said user automatically becomes the first admin within that organization. A description of the roles and their corresponding privileges is outlined as follows:

- Member: Individuals of this role have the ability to view the private dataset of the organization.
- Editor: In addition to the privileges afforded to members, editors have the ability to add new datasets, modify existing ones and also manage their visibility status.
- Admin: the role that holds the highest authority. Admins can perform all editor functions, and in addition, add or remove users, adjust user roles (including other admins users), refine and edit the details of the organizations (title, description, image etc.) and, if necessary, even delete the organization entirely.

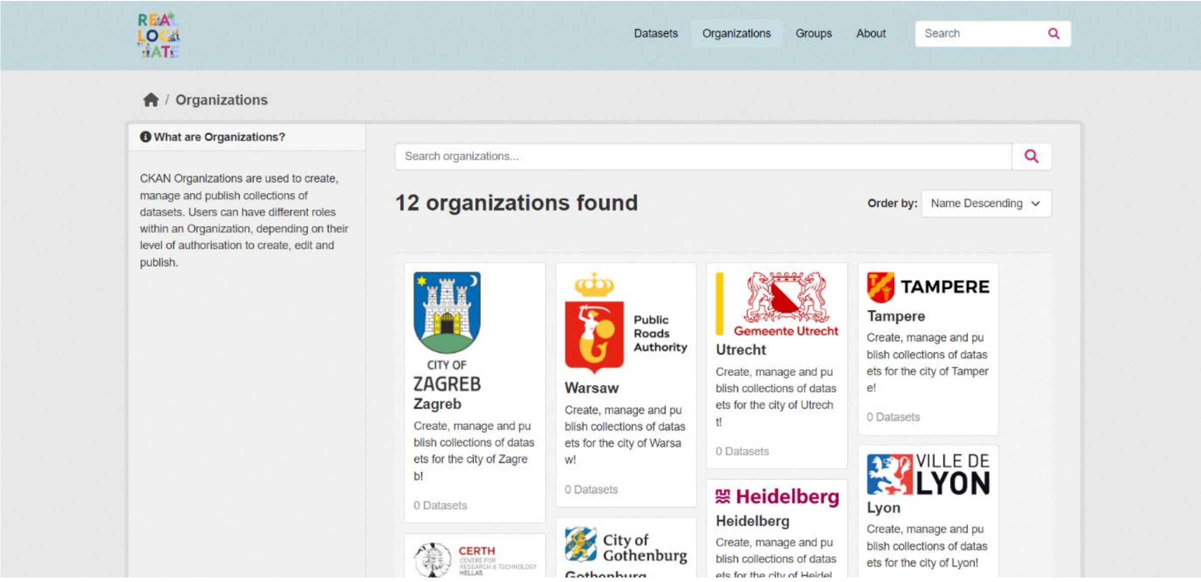

Figure 6. Organizations in REALLOCATE's data repository.

4.4.3 Datasets and their properties

Datasets serve as the fundamental unit of data dissemination and collaboration within CKAN. They represent individual collections of data, such as databases, documents, spreadsheets or any other form of structured information. The dataset page of the REALLOCATE data repository can be seen in Figure 7. The datasets featured therein are currently provided for testing and visualisation purposes and are prone to modifications in future iterations.

The "Datasets" page is divided into two primary segments boosting clarity and ease of navigation. The left column, as seen in Figure 7, houses the information section, where users can explore and see all existing organizations, group classifications, data tags and format specifications inside the repository, and even potential licenses that could be applicable to the data. On the other hand, the datasets section provides a catalogue of all available datasets paired with their description.

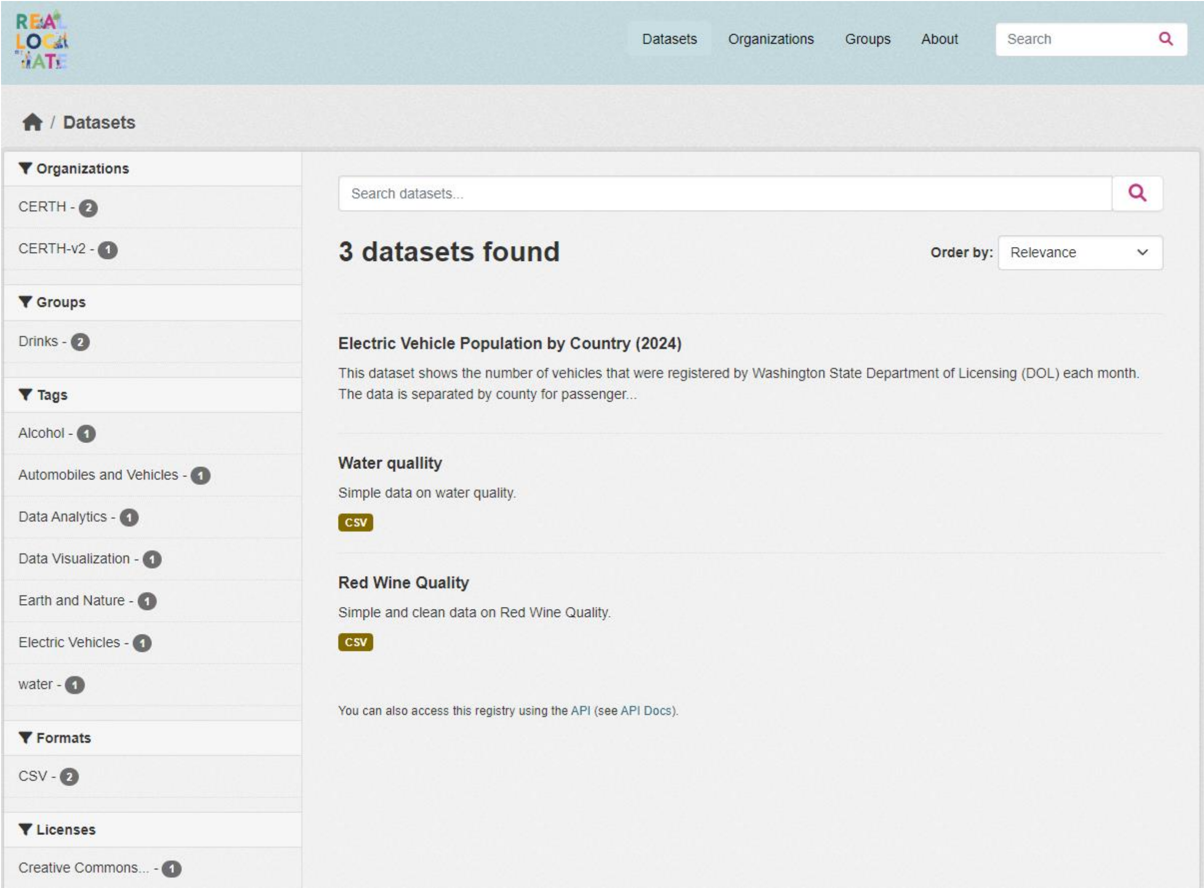

Figure 7. Dataset page in REALLOCATE's data repository.

In the same page, users also have the ability to add new datasets. When attempting to add a new dataset, users are presented with the intuitive "Create Dataset" form, as seen in Figure 8, where details and information regarding the data will be asked. Specifically, the information fields that users are prompted to add are as follows:

- Title: A comprehensive and yet specific title, serving as a unique identifier across the CKAN platform.
- Description: A longer description of the dataset, including insights such as origin and additional details necessary for users to effectively utilize the data.
- Tags: Users can add tags to aid and enhance discoverability of the data. Tags are also used to establish connections with related datasets.
- License: A drop-down box with license information to classify your data. It is important to include this information so users know if they are eligible to use the data.
- Organization: If a user is affiliated with multiple organizations, this drop-down box allows users to specify under which organization will this specific dataset be owned.

- Visibility: Either public or private. Public datasets are accessible to all site users, whereas private datasets restrict access solely to members of the owning organization, while remaining hidden from the general searches.
- Source: If applicable, a Uniform Resource Locator (URL) directed to the source of the data.
- Version: A numeric indication specifying the version of the data, particularly beneficial in case multiple instances are created.
- Author/Author email: The name of the individual, or organization, responsible for generating the dataset, along with an email address serving as a point of contact for queries related to the data.
- Maintainer/Maintainer email: If applicable, a secondary individual responsible for the data.

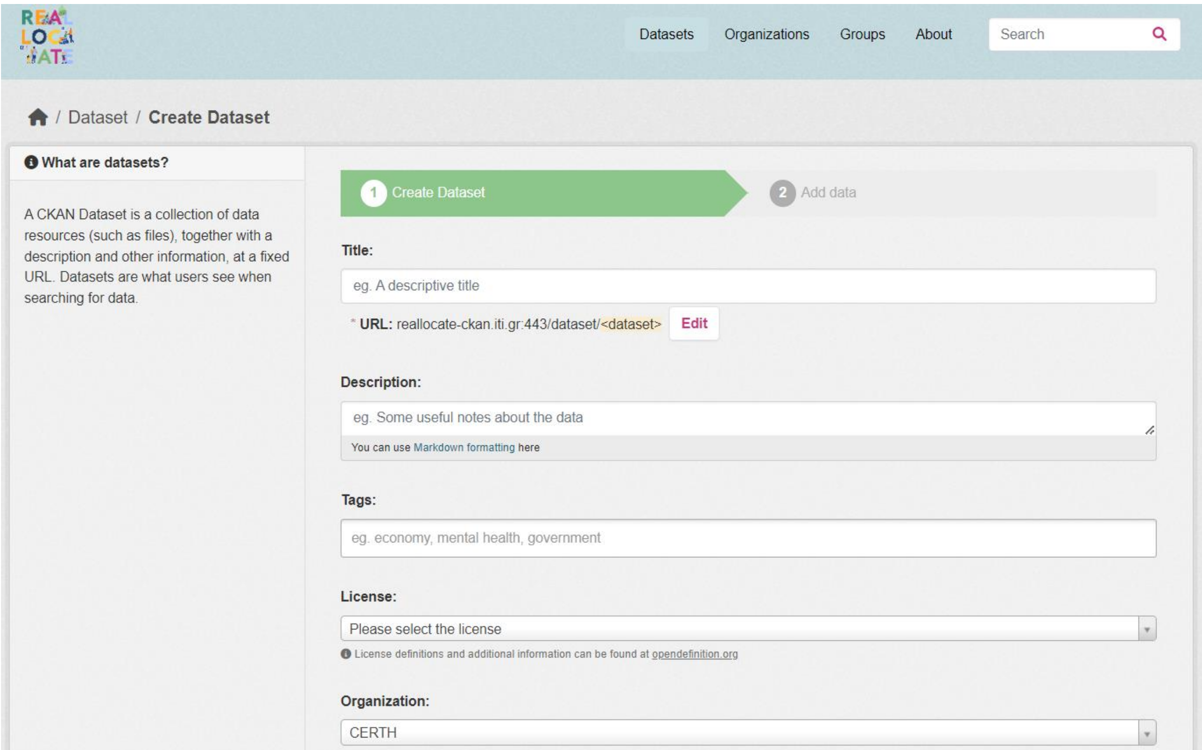

Figure 8. "Create Dataset" form in REALLOCATE's data repository.

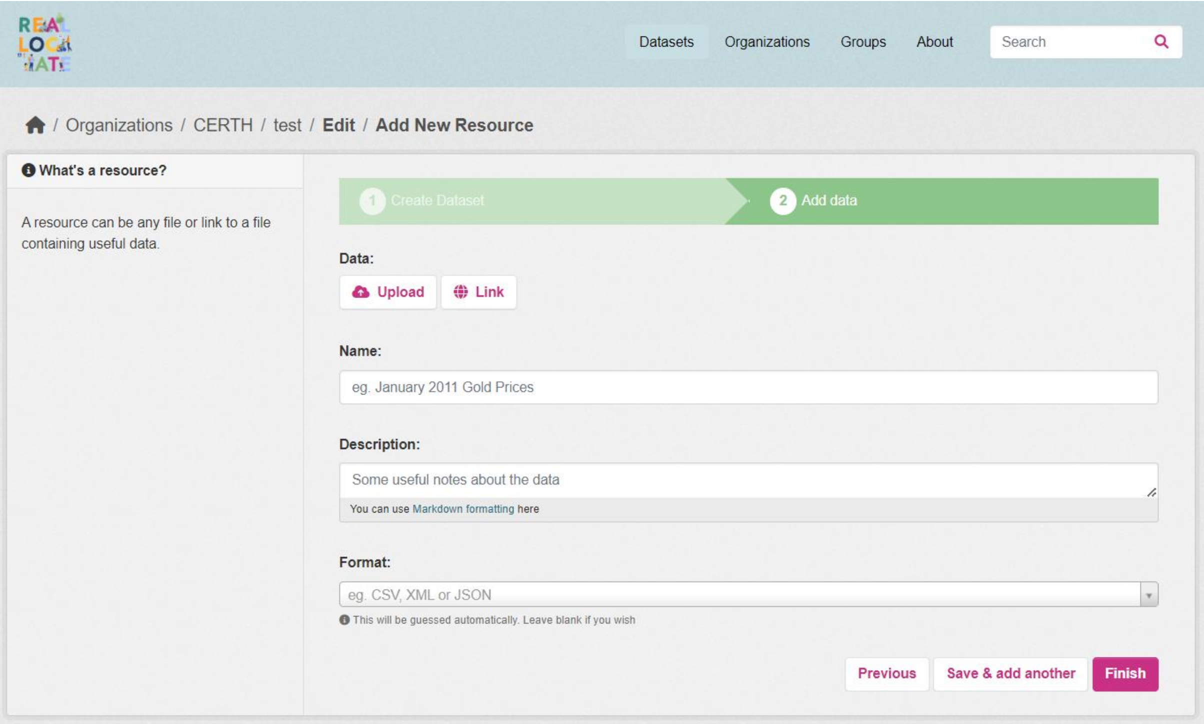

Figure 9. "Add data" form in REALLOCATE's data repository.

It is important to mention that among the fields required, only the "Title" field stands as mandatory, while all other information remain optional. However, it is highly advisable to include as many information as possible to ensure the appropriate use of the respective dataset. After specifying the aforementioned details, users progress to the subsequent and final stage, denoted as the "Add Data" form, as presented in Figure 9. Within this form, users have the option to either upload the data directly from their device, or specify a link pointing to the dataset. Optionally, a name, description and file format can be also specified for this very instance of the data. Notably, the information within both forms retains the flexibility to be modified in the future.

4.4.4 Groups

Groups within CKAN are a vital mechanism used to cluster related datasets together for easier discovery and accessibility. They are primarily used to categorise and organise datasets based on common themes, topics or attributes they posse. An example of groups can be seen in Figure 10, where both two existing datasets of REALLOCATE repository are included in the group "Drinks". It is important to make a distinction between groups and organizations, since these two terms have some key differences within CKAN. Specifically, groups are only used to cluster datasets together without any membership or role attribution.

Datasets associated with groups, may still be subject to their individual access controls, particularly if they are published through an organization with private visibility, limiting access of the dataset to only members of that particular organization. Finally, a dataset can belong to a singular organization, while the same dataset can be associated with multiple groups.

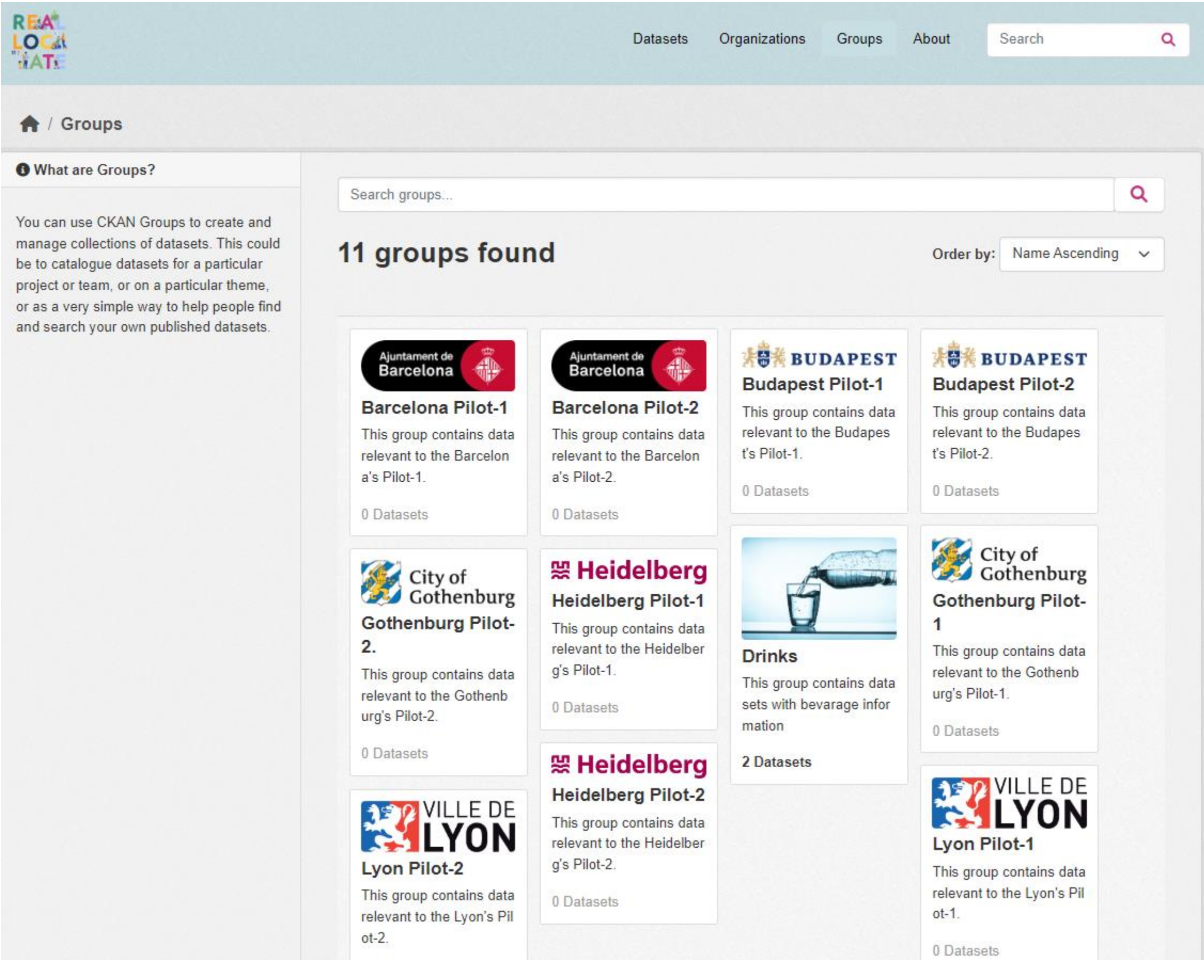

Figure 10. Groups in REALLOCATE's data repository.

4.4.5 Visualisation of datasets

CKAN extends its capabilities beyond mere data storage by offering visualisation options for the uploaded datasets. Via the resource preview page, users can access various visualisation formats, including grid tables, graphs and map views (Figure 11). These visualisations are commonly referred to as resource views, and represent the default functionalities provided by CKAN. Resource views typically have the following key features:

- One resource (dataset) can have multiple views simultaneously (for example both grid table and a graph for the same data).
- Users with editor privileges can reorder and configure these views individually, and also select which will be visible.
- Views can be embedded on external sites.

In addition to the default views, CKAN offers the option to activate supplementary plugins, thereby expanding the available visualisation options. Once custom plugins are added, administrators are able to define and choose the default views for specific resource formats, enhancing user experience and efficiency. Some examples of additional view plugins include:

- Dashboard: A combination of multiple views integrated in the same interface.
- PDF viewer: Ability to render PDF files directly on the resource page.
- GeoJSON map: Ability to render GeoJSON files on an interactive map interface.
- Choropleth map: Display statistical geographical data through a choropleth map visualisation.
- Basic charts: Offers diverse graph types and rendering options.

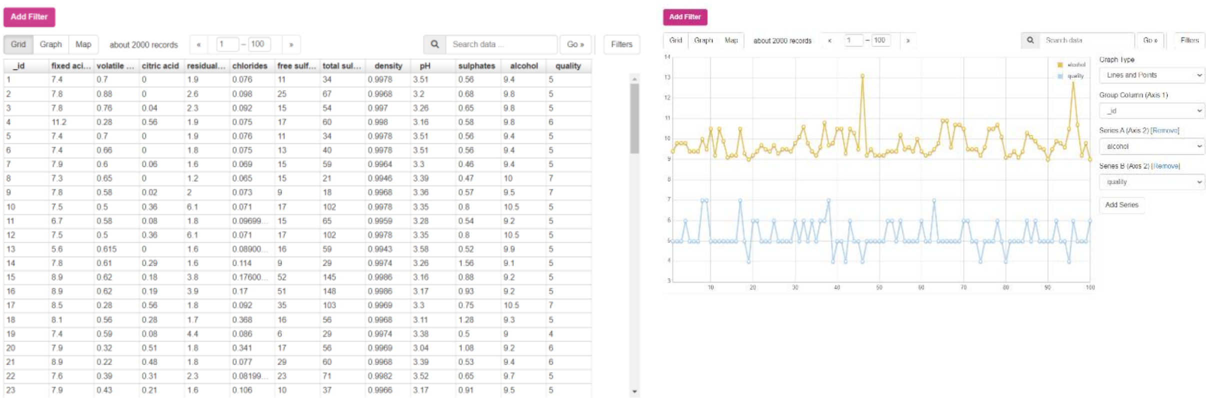

Figure 11. View examples in REALLOCATE's data repository.

5 AI-powered Distributed Dashboard

This section describes the role of the AI-powered Distributed Dashboard within REALLOCATE, detailing its architectural framework and the technologies required for its development. Finally, it presents the primary interface and the critical information for each participating pilot site that will be displayed on the dashboard.

5.1 Role overview

The REALLOCATE project's AI-powered Distributed Dashboard is a custom-designed platform tailored to visualise KPIs from various collaborating test sites. Its primary purpose is project monitoring, serving as a centralised interface for accessing processed data, research outputs, and tools/services developed within the project's SSMLs. This dashboard embodies flexibility, interoperability, and user-friendliness, serving as an open entry point for users.

The REALLOCATE AI-powered Distributed Dashboard represents a sophisticated platform, serving as the primary interface for end-users accessing a diverse array of services, including data analytics, visualisations, and monitoring systems. The main objective is to establish a dynamic environment conducive to augmented analytics, thereby enhancing the management and evaluation of interventions in the pilot cities involved. Through standardization efforts, the dashboard aims to ensure equitable proficiency and capabilities across all participants within the platform ecosystem. Ultimately, the REALLOCATE dashboard seeks to revolutionize urban data management and decision-making processes.

The subsequent focus of this chapter shifts towards the architectural framework underpinning the REALLOCATE dashboard and its specific components. It will describe the workflow from data collection across all participating cities, through aggregation and processing within the data management platform, concluding in visualisation on the dashboard interface. This comprehensive exploration aims to explain the detailed workings of the REALLOCATE dashboard and its pivotal role in facilitating informed decision-making for sustainable urban development.

5.2 Justification for a Centralised Dashboard

Given the diversity and varying levels of existing digital infrastructure across participating cities, REALLOCATE has opted to develop a centralised, comprehensive dashboard rather than upgrading individual city dashboards. This decision ensures a standardised approach to KPI monitoring and data visualisation, providing equitable access to advanced analytics and decision-support tools for all cities, regardless of their current technological capabilities.

Out of the 10 cities participating in the project, only Tampere and Utrecht currently have publicly accessible dashboards, as shown in Figure 12 and Figure 13. However, the REALLOCATE dashboard serves different needs than these existing dashboards, which focus mostly on car and traffic data. Regardless, the insights from the existing dashboards can be embedded in the unified dashboard with website embeddings. Most of the cities already have open data repositories but lack a way to visualise their data. By using the existing APIs to access open data from these cities, the REALLOCATE dashboard can incorporate these data, enhance visualisations, and provide more comprehensive insights.

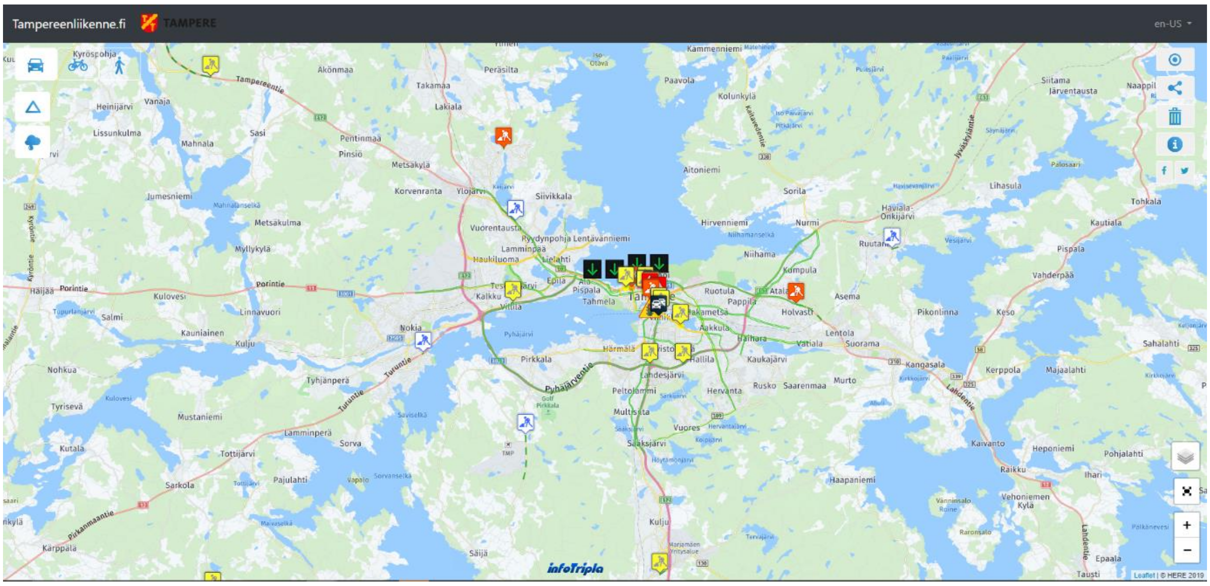

Figure 12. Tampere traffic dashboard

This centralised dashboard serves as a unified platform where users can navigate to each pilot site and monitor relevant KPIs. While some cities may already have their own dashboards, these will be integrated into the main comprehensive dashboard where feasible. This integration ensures that the comprehensive dashboard fulfils all the requirements of REALLOCATE, delivering a coherent and consistent user experience across all pilot sites. This approach not only simplifies the monitoring process but also enhances collaboration and data sharing among cities, aligning with the project's goal of fostering inclusiveness. For cities without any previous solution, a new dashboard will be provided to interact with the rest of the platform, ensuring a cohesive and integrated approach to urban data management.

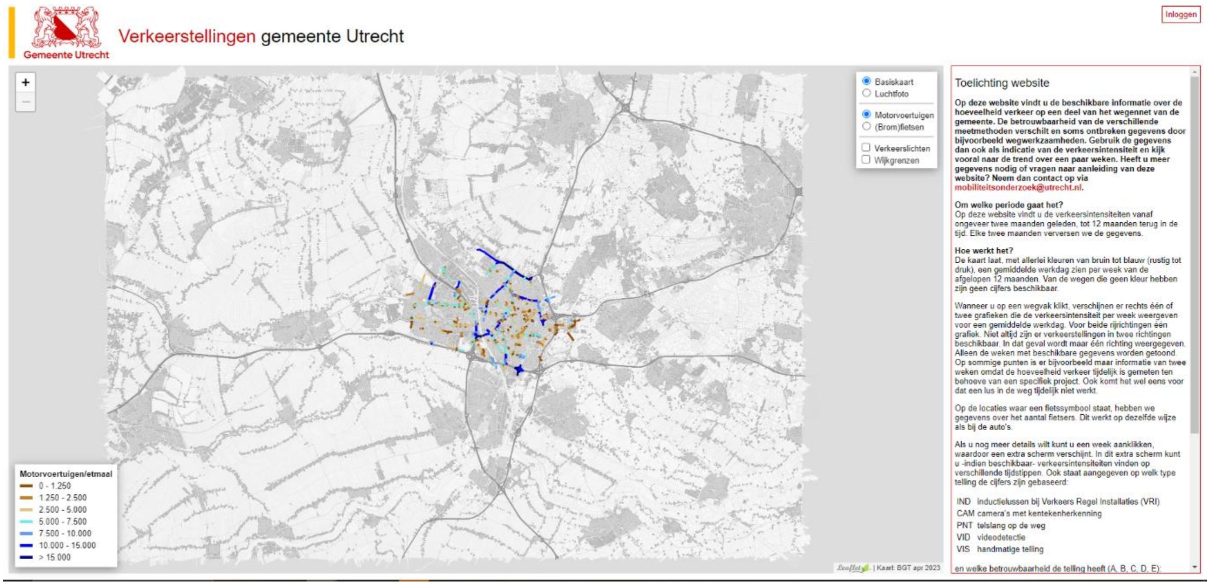

Figure 13. Utrecht traffic dashboard

5.3 Architecture

To effectively meet the needs of REALLOCATE and deliver a robust and resilient AI-powered distributed dashboard, a well-defined architecture and design principles are essential. This architecture, illustrated in Figure 2 consists of three key components: the cities, the data management platform, and the dashboard interface. Each part plays a crucial role in shaping the final solution, ensuring that it aligns with the project's vision of transforming urban spaces into inclusive, green, safe, and future-proof environments where communities can thrive. A detailed description of all three parts will be given in the following sections, outlining how each part of the architecture contributed for the creation of the final solution.

5.3.1 Cities

In the context of REALLOCATE, the pilot sites participating in the initiative encompass cities designated as both lead and twin cities within the Climate-Neutral & Smart Cities EU mission. The primary objective for each pilot site, as outlined in the design principles of Figure 2, is to maximise the utilisation of existing infrastructure and implement necessary updates to facilitate seamless data collection. This may involve conducting online surveys, sharing web service feeds, and gathering historical data to enhance data accessibility. Additionally, each pilot site has the autonomy to implement tailored actions based on specific criteria, such as:

- Improving the collection of pedestrian, bike, and e-scooter traffic data.

- Establishing reporting mechanisms for incidents involving pedestrians, cyclists, and e-scooter users.
- Developing and executing strategies to assess accessibility for individuals with disabilities in shared spaces.

The historical data collected plays a crucial role in generating valuable insights for the second part of the REALLOCATE dashboard architecture. In addition to historical data, each pilot site has the capability to provide pre-calculated KPIs for direct integration into the corresponding section of the REALLOCATE dashboard. This comprehensive approach ensures that the dashboard not only presents historical trends but also offers real-time performance metrics, enabling stakeholders to make informed decisions and drive sustainable urban development initiatives effectively.

5.3.2 Data management platform

At the core of the REALLOCATE dashboard architecture lies a robust data management platform, designed to securely store, process, and transform the vast amounts of data collected from the pilot sites. This platform is the backbone of the system, ensuring that all data is accurately and efficiently utilised to generate valuable insights and inform decision-making. By harnessing the power of AI-powered solutions, the data management platform can calculate the required KPIs, which are then displayed on the dashboard, providing stakeholders with a comprehensive view of the project's progress.

The data management platform begins by authenticating pilot sites, allowing them to securely upload their data. This ensures that only authorised entities have access to the platform, safeguarding the integrity of the data and the project as a whole. Once data is uploaded, the platform communicates with the database to facilitate data access. The data is then converted into a common format, compatible with the AI-powered methods used for analysis. This conversion process is critical, as it enables the seamless integration of data from diverse sources, ensuring that all data is uniformly processed and analysed.

5.3.2.1 AI-Powered Calculation of KPIs

The data management platform leverages AI-powered solutions to calculate the KPIs. These KPIs are essential for measuring the success of the REALLOCATE project, providing stakeholders with a clear understanding of the project's impact and progress. The platform's ability to accurately calculate KPIs is a testament to its robustness and effectiveness in managing the vast amounts of data collected.

In this subsection, examples of how the KPIs will be calculated based on the pilot sites' input are described. This approach is based on the work done in Task 5.1 and the input from the cities so far. Therefore, there may be changes in the future according to updates from the cities. Table 2 illustrates some of the KPIs definitions, the indicator names and definitions which describe how the KPIs will be calculated, the data sources used for collecting the data, the frequency at which this data will be collected, and the final display on the dashboard. In the final phase, similar procedures will be followed for the complete set of KPIs.

Table 2. KPIs calculation methods

KPI	Indicator name	Indicator Definition	Data source	Frequency	Dashboard display
Quality of public spaces	Incident reduction	Track the number of accidents, near misses, or safety incidents reported at intersections before and after safety auditing procedures.	Police traffic incident reports, traffic cameras, speed monitoring equipment	Quarterly	Number of incidents
Reallocation of public space (sqm/year)	Bike parking capacity	The increase in the number of bike parking spaces or bike-sharing docks installed as part of the project.	City infrastructure records, site surveys	Semi-annually	Number of feedback responses
Cycling & e-	Increased	The percentage	City	Annually	Percentage

bike comfort	Bike Parking Spaces	increase in the number of dedicated bike parking spaces in the targeted areas.	infrastructure records, site surveys		(%) from baseline
Pedestrian & disabled comfort	Improved Walkable Conditions	The percentage increase in the quality of walkable conditions, including sidewalk conditions, accessibility features, and pedestrian amenities, compared to a baseline measurement before the project's implementation.	Walkability audits, infrastructure surveys	Annually	Percentage
Increases in pedestrians + cyclists	Cyclist Count	The total count of cyclists observed or recorded in the area.	Traffic sensors, manual counts	Daily/Hourly	Count
Traffic safety active modes	Community Engagement Index	The level of community engagement and	Surveys, interviews, engagement logs	Quarterly	Index/Rating (1-100)

		acceptance in the co-creation process and satisfaction with the implemented safety measures.			

The KPIs will be updated dynamically and displayed on the dashboard. Users will be able to compare the KPIs between baseline and intervention conditions. Additionally, the KPIs will be stored in the database to enable export if necessary. As the project progresses, additional features that could enhance the interpretation and comparison of KPIs will be explored and included in future versions.

5.3.2.2 Integration with CKAN

The REALLOCATE data management platform is built upon the CKAN framework, a widely recognized open-source data management system. CKAN provides a robust infrastructure for managing and sharing data, ensuring that the data is secure, accessible, and easily searchable. By integrating with CKAN, the REALLOCATE platform benefits from a mature and widely adopted data management solution, allowing for the efficient management of data and the seamless integration of data from diverse sources.

5.3.3 Keycloak

Keycloak[28] is an open-source solution for web applications and services, designed to facilitate single sign-on functionalities. Positioned at the forefront of modern security infrastructure, Keycloak aims to simplify and streamline security processes, allowing developers to leverage out-of-the-box security features tailored to suit the needs of any organization. Some of the key features provided by Keycloak include:

- Single sign-on and single sign-out for browser applications
- Support for industry-standard protocols such as OpenID Connect, OAuth 2.0, and SAML

[28] https://www.keycloak.org/docs/latest/server_admin/

- Social login integration with platforms like Google, Facebook, and GitHub
- Flexible customisation options
- Advanced security measures, including two-factor authentication and login flows (self-registration, password recovery, and email verification)

Keycloak functions as an independent server, through which applications are configured to establish connections. When a user accesses browser applications, a redirection to the Keycloak authentication server occurs, and credential input is required. This redirection is crucial, as it ensures that the application itself will exclusively receive only cryptographically signed identity tokens from users, thereby maintaining complete user isolation from the underlying applications.

By leveraging Keycloak's robust security features and streamlined authentication processes, developers can focus on building their applications without the burden of implementing complex security measures from scratch. This allows for the creation of secure and user-friendly web applications that seamlessly integrate with Keycloak's single sign-on capabilities.

5.3.4 NGINX

Nginx[29] is a high-performance web server and reverse proxy that is widely used in the industry for its efficiency and scalability. As a key component in the REALLOCATE project's architecture, Nginx plays a crucial role in ensuring reliable and secure communication between clients and the backend services. Some of the key features and functionalities of Nginx include:

- Efficient handling of concurrent connections and high traffic loads
- Load balancing capabilities to distribute incoming traffic across multiple servers
- SSL/TLS termination for secure communication over HTTPS
- Web server functionalities for serving static content and handling dynamic requests
- Reverse proxy capabilities for routing requests to backend services

By incorporating Nginx into the REALLOCATE architecture, the project benefits from improved performance, enhanced security, and seamless scalability. Nginx acts as a robust intermediary between clients and the backend services, optimizing the delivery of data and ensuring a smooth user experience. Its versatility and reliability make it a valuable asset in

[29] https://nginx.org/en/

managing the flow of data and requests within the distributed dashboard system, contributing to the overall efficiency and responsiveness of the platform.

5.3.5 MongoDB

MongoDB[30] is an open-source NoSQL database that serves as the primary storage solution for all historical data accessible through the API handlers of CKAN. Due to its capability to manage large volumes of data without strict schema requirements, MongoDB is the optimal choice for the needs of REALLOCATE. Some of the key features and advantages of MongoDB include:

- Flexible Schema[31]: MongoDB offers a dynamic schema, allowing data to be stored without a predefined structure. This flexibility promotes agility and responsiveness, enabling quick adjustments to accommodate new requirements or changes in data structure within REALLOCATE.
- Scalability[32]: MongoDB excels in scalability, handling increasing data, traffic, and workload efficiently without compromising performance. It achieves scalability through vertical scaling (increasing resources of a single server) and horizontal scaling (distributing workload across multiple servers via sharding, where data is partitioned across nodes).
- High Performance[33]: MongoDB's architecture is optimized for high performance across various operational demands. Features like automatic indexing, in-memory caching, and native support for replication and load balancing always ensure efficient data access.
- Replication[34] and Fault Tolerance[35]: MongoDB employs a sophisticated data replication strategy across multiple nodes, enhancing fault tolerance and ensuring data integrity. In the event of hardware failures or network issues, MongoDB automatically performs failover to a secondary replica, minimising downtime and data loss.

By leveraging MongoDB's robust features, REALLOCATE benefits from a reliable and efficient data management system that can handle the diverse and evolving needs of the

[30] https://docs.mongodb.com/
[31] https://www.mongodb.com/docs/manual/core/data-model-design/
[32] https://www.mongodb.com/docs/manual/core/sharding/
[33] https://www.mongodb.com/docs/manual/core/index-types/
[34] https://www.mongodb.com/docs/manual/core/replica-set-high-availability/
[35] https://www.mongodb.com/docs/manual/replication/

project. MongoDB's scalability, performance optimization, and fault-tolerant design make it an asset in ensuring the seamless operation and integrity of the distributed dashboard architecture.

5.4 Visualisation

At the forefront of the REALLOCATE project's architecture is the dashboard, serving as the primary interface that consolidates and displays critical information for each participating pilot site. The fundamental purpose of this dashboard is to provide a visual snapshot of the performance achieved by each site, presenting the respective KPIs that are relevant to the project's objectives.

By acting as an effective management and strategic planning platform, the dashboard facilitates the delivery of useful information, enabling quick, informed, and impactful decision-making. A visual representation of the dashboard can be seen in Figure 14, which comprises three primary components: the left sidebar, the map, and the KPIs field.

5.4.1 The Left Sidebar: Navigating the Dashboard

The left sidebar of the dashboard serves as the gateway for users to access and explore the various features and functionalities of the platform (Figure 15). This intuitive navigation menu allows stakeholders to seamlessly transition between different views, data visualisations, and analytical tools, empowering them to delve deeper into the insights and data relevant to their specific roles and responsibilities.

5.4.2 The Map: Visualising Spatial Data

The map component of the dashboard provides a powerful spatial representation of the data, enabling users to visualise the performance and progress of the REALLOCATE project across the participating pilot sites. By overlaying various data layers, such as transportation patterns, infrastructure updates, and community engagement metrics, the map offers a comprehensive and interactive view of the urban transformation taking place in each city (Figure 16).

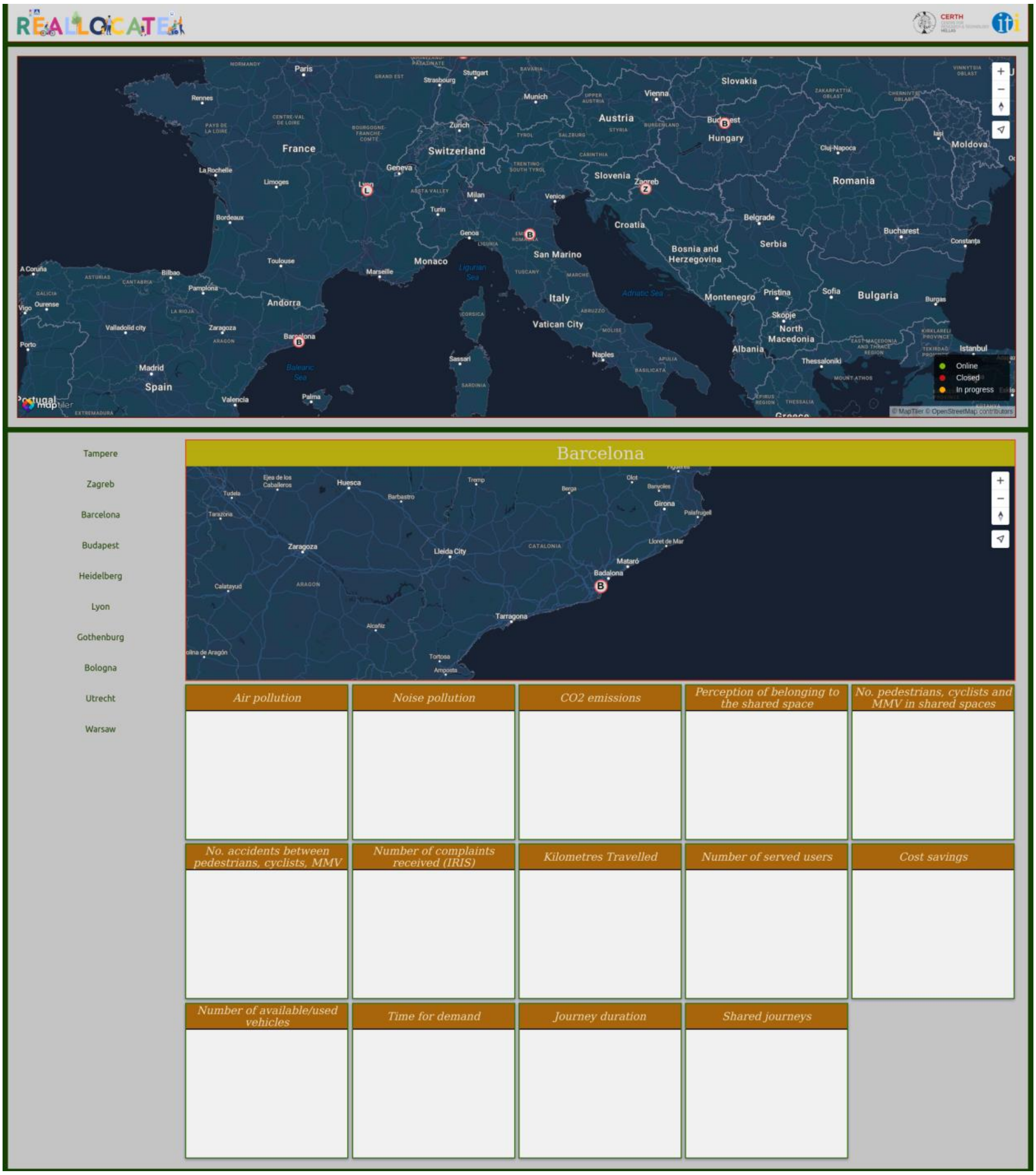

Figure 14. The REALLOCATE Dashboard.

Figure 15. Left sidebar of REALLOCATE Dashboard.

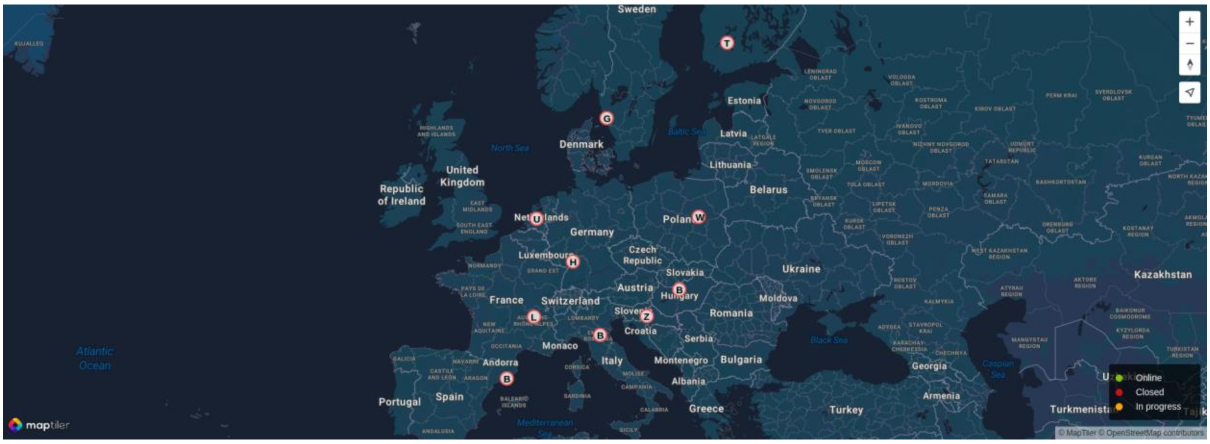

Figure 16. The map of REALLOCATE Dashboard.

5.4.3 The KPI Field: Tracking Performance

The KPIs field is the centerpiece of the dashboard, presenting the key performance indicators that are crucial for measuring the success of the REALLOCATE project (Figure 17). This section offers a concise and visually appealing display of the relevant KPIs, allowing stakeholders to quickly assess the progress made by each pilot site and identify areas that require further attention or intervention.

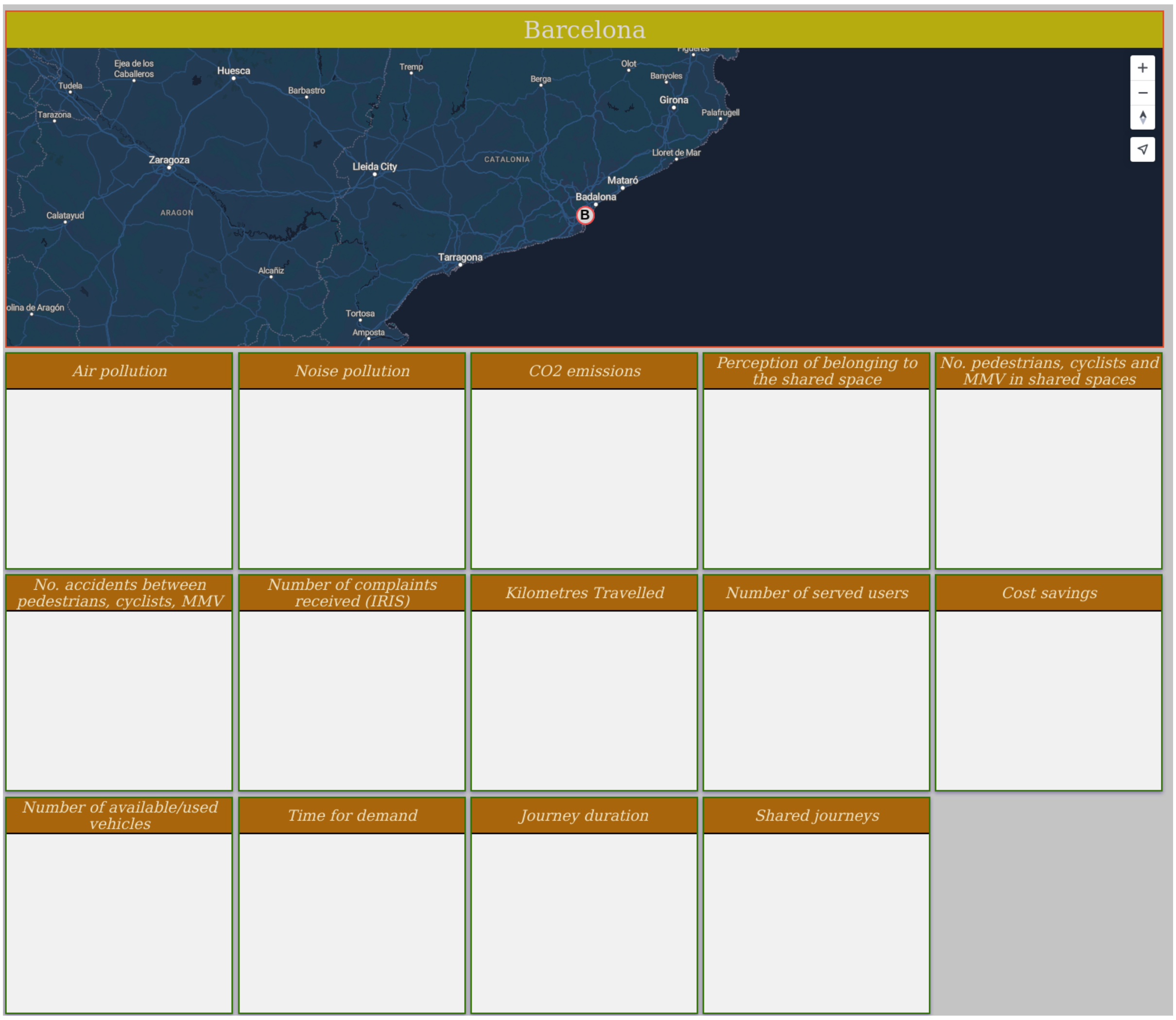

Figure 17. KPI Field of REALOCATE Dashboard.

6 Future work and conclusions

The current deliverable has successfully outlined the progress in Task 5.2, the AI-powered Distributed Dashboard and the Multimodal Data Hub, essential for the continuous monitoring of KPIs. This document has detailed the architectural framework consisting of three main components: a) the pilot sites, b) the data management platform and c) the dashboard.

The multimodal data hub has proven pivotal in storing and managing diverse data types across cities, which is instrumental for calculating and monitoring KPIs in real-time. The deployment of CKAN as the underlying data management platform supports these functionalities by providing a robust, secure, and flexible environment for data interaction. Additionally, the development of both the back-end and front-end parts of the dashboard has been presented in detail.

As the project moves forward, several enhancements are planned to refine the system's efficiency and responsiveness. The next steps include:

1. **Implementing Additional Customisations:** Tailoring the dashboard to better meet the specific needs and requirements of REALLOCATE, ensuring it remains adaptable and user-friendly for all stakeholders involved.
2. **Standardising Data Formats:** Establishing uniform data formats through advanced ETL processes to facilitate smoother data integration and analysis across the platform.
3. **KPIs calculation:** Integrating procedures that take as input the raw data uploaded from the cities to the data management platform and extracting the final KPIs.
4. **Enhancing Real-Time Data Processing:** Integrating more sophisticated real-time data processing capabilities to provide up-to-the-minute updates on KPIs, thereby enabling more dynamic decision-making.
5. **Expanding Visualisation Tools:** Further developing and integrating advanced visualisation tools such as dynamic charts, plots, and diagrams to aid in the more effective representation and understanding of cities data.

References

[1] Y. Karimi, M. Haghi Kashani, M. Akbari and E. Mahdipour, “Leveraging big data in smart cities: A systematic review,” *Concurrency and Computation: Practice and Experience,* vol. 33, June 2021.

[2] J. Pereira., S. Sargento. and J. M. Fernandes., “Decision Support Dashboard for Traffic and Environment Analysis of a Smart City,” in *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems - VEHITS*, 2018.

[3] R. Gubareva and R. P. Lopes, “Literature Review on the Smart City Resources Analysis with Big Data Methodologies,” *SN Computer Science,* 2024.

[4] A. Joyce and V. Javidroozi, “Smart city development: Data sharing vs. data protection legislations,” *Cities,* vol. 148, p. 104859, 2024.

[5] J. Macedo, N. Cacho and F. Lopes, “A Comparative Study of Tools for Smart Cities Open Data Publication and Management,” in *2017 IEEE First Summer School on Smart Cities (S3C)*, 2017.

[6] X. Liu, A. Heller and P. S. Nielsen, “CITIESData: a smart city data management framework,” *Knowledge and Information Systems,* 2017.

[7] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan and R. Maciejewski, “VAUD: A Visual Analysis Approach for Exploring Spatio-Temporal Urban Data,” *IEEE Transactions on Visualization and Computer Graphics,* vol. 24, pp. 2636-2648, 2018.

[8] G. Suciu, L. Necula, T. Usurelu, I. Rogojanu, M.-C. Diţu and A. Vulpe, “IoT-Based 3D Visualisation Platform for an Efficient Management of the Smart City Ecosystem,” in *2018 Global Wireless Summit (GWS)*, 2018.

[9] S. S. Sergio Trilles and J. Huerta, “A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data,” *International Journal of Digital Earth,* vol. 10, p. 103–120, 2017.

[10] A. S. Alic, J. Almeida, G. Aloisio, N. Andrade, N. Antunes, D. Ardagna, R. M. Badia, T. Basso, I. Blanquer, T. Braz, A. Brito, D. Elia, S. Fiore, D. Guedes, M. Lattuada, D. Lezzi, M. Maciel, W. Meira, D. Mestre, R. Moraes, F. Morais, C. E. Pires, N. P. Kozievitch, W. dos Santos, P. Silva and M. Vieira, "BIGSEA: A Big Data analytics platform for public transportation information," *Future Generation Computer Systems,* vol. 96, pp. 243-269, 2019.

[11] S. Fiore, D. Elia, C. E. Pires, D. G. Mestre, C. Cappiello, M. Vitali, N. Andrade, T. Braz, D. Lezzi, R. Moraes, T. Basso, N. P. Kozievitch, K. V. O. Fonseca, N. Antunes, M. Vieira, C. Palazzo, I. Blanquer, W. Meira and G. Aloisio, "An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management," *IEEE Access,* vol. 7, pp. 117652-117677, 2019.

[12] D. Lahat, T. Adali and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE,* vol. 103, p. 1449–1477, 2015.

[13] G. Trappolini, A. Santilli, E. Rodolà, A. Halevy and F. Silvestri, "Multimodal neural databases," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023.

[14] M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas and others, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data,* vol. 3, 2016.

[15] N. Lopes, M. Stephenson, V. Lopez, P. Tommasi and P. M. Aonghusa, "On-demand integration and linking of open data information," in *Research Conference on Metadata and Semantics Research*, 2015.