Data Management Plan

Deliverable number: D1.4

Date: 31/10/2024

Version: v2

Name of work package	WP1 – Project coordination
Dissemination level	PUB
Due delivery date:	31/10/2024
Actual delivery date:	31/10/2024
Lead beneficiary:	CERTH/ITI
Contributing beneficiaries:	-
Authors:	CERTH/ITI: Theoktisti Marinopoulou, Antonios Lalas, Konstantinos Votis
Peer reviewers:	CERTH/HIT: Katerina Touliou, UCD: Aisling Jackman

Document history				
Version	**Submitted for review by**	**Date**	**Reviewed by**	**Date**
1	Theoktisti Marinopoulou	18/10/24	Katerina Touliou Aisling Jackman	21/10/24 25/10/24
2	Theoktisti Marinopoulou	14/02/25	Katerina Deliali (PO; CINEA)	15/01/25

Executive Summary

The current deliverable is designed to establish an updated Data Management Plan (DMP) aligning with the European Commission's Findable, Accessible, Interoperable, Reusable (FAIR) principles. It aims to ensure comprehensive data management and General Data Protection Regulation (GDPR) compliance through refined data security, pseudonymization, and enhanced data-sharing methods. This deliverable has introduced enhancements in data collection, storage, and protection practices. The plan now includes a secure repository for pilot data, guidelines on metadata provision, and rigorous data accessibility protocols that support transparency and compliance across the project's lifecycle.

D1.4 constitutes the second version of the DMP for the REALLOCATE project. The first version, D1.3, was submitted in M6 and a subsequent version, D1.5, is scheduled to be released in M36. This second iteration of the DMP, introduces several refinements and updates to the data management strategy. The deliverable is directly associated with Task 1.6, which focuses on the creation, updating, and management of the DMP throughout the project lifecycle. It is also closely linked to several key tasks, particularly those involving data handling, storage, and processing.

More specifically, this deliverable outlines the comprehensive data management strategy of the project. It details the approach towards data collection, curation, storage, preservation, and sharing, while ensuring compliance with the European GDPR and associated national legislations. Adhering to the FAIR data management principles, the deliverable is focusing on ensuring data findability, accessibility, interoperability, and reusability.

Additionally, it describes the project's data repository which constitutes a shared, secure space where all pilot data can be uploaded. Moreover, the plan highlights measures for data security and protection, emphasizing the importance of pseudonymisation and anonymisation in data processing.

This document will serve as a reference point for all consortium members, ensuring the proper and ethical management of data throughout the project's lifecycle.

Table of Contents

REALLOCATE Deliverable 1.4 Data Management Plan

List of Tables

List of Figures

REALLOCATE Deliverable 1.4 Data Management Plan

List of Acronyms

Acronym	Explanation
AI	Artificial Intelligence
API	Application Programming Interface
CA	Consortium Agreement
CKAN	Comprehensive Knowledge Archive Network
CSV	Comma-Separated values
DMP	Data Management Plan
DMZ	Demilitarized Zone
DOI	Digital Object Identifier
DRA	DOI Registration Agency
EC	European Commission
ETSI	European Telecommunications Standards Institute
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IPR	Intellectual Property Rights

KPI	Key Performance Indicator
PC	Project Coordinator
PDF	Portable Document Format
SotA	State of the Art
SQL	Structured Query Language
SRC	Source Code Repository
SSL	Secure Sockets Layer
SSML	Safe & Sustainable Mobility Labs
VRU	Vulnerable Road User

1 Introduction

This deliverable is designed to detail the management, processing, and protection of all data types including pilot data, scientific data and any data generated or curated throughout the REALLOCATE project. Establishing a comprehensive Data Management Plan (DMP) is essential in aligning with the European Commission's stipulations and best practices for responsible data management. It's vital to emphasize that the DMP is a dynamic document. As the project progresses and as research and development objectives may shift, the final version may undergo further refinement. The types of outcomes expected from the REALLOCATE project include:

• **Project Deliverables:** Detailed documentation on various milestones and results achieved during the REALLOCATE project.

• **Scientific Publications:** Research papers, journal articles, and other scholarly outputs that derive from the project's findings and methodologies.

• **Contribution to Standards:** Inputs to define or refine standards, especially those related to Findable, Accessible, Interoperable and Reusable (FAIR) data management principles, data security, and privacy.

• **Software and Applications:** Digital tools, applications, and interfaces, such as the REALLOCATE Distributed Dashboard and Multimodal Data Hub, aimed at facilitating data navigation, analytics, visualization, and other functionalities.

• **Data Collected for Analysis and Evaluation:** This pertains to all the data sets acquired for the project, whether through sensors, reports, surveys, or other means, and post-processing steps, including pseudonymization, anonymization, and other GDPR-compliant measures.

• **Pilot Results Datasets:** Datasets specifically stemming from the 10 Safe & Sustainable Mobility Labs (SSMLs) to the REALLOCATE project, inclusive of traffic, safety, environmental, climatic, behavioral and mobility intervention data.

1.2 Deliverable Context

The DMP will shape the way the data and associated information are managed, aligning with the undermentioned objectives:

- Reinforce and supplement the goals of EU directives and guidelines.

- Compliance with legal, ethical, privacy and personal data regulations and fundamental rights.
- Widespread dissemination efforts and robust public engagement.
- Establishment of a widespread network of stakeholders and collaboration with security projects and related initiatives.

The deliverable D1.4 – Data Management Plan represents the updated version of the DMP. The initial version, D1.3, was submitted at M6, and the final version, D1.5, is scheduled for submission at M36. As the project progresses, any data-related considerations should be viewed through the prism of this DMP.

The D1.3 was linked with the first milestone of the project which is "First version of Data Management Plan".

The DMP is designed to responsibly oversee and protect all data and its accompanying metadata that generated from the project's research, evaluations, or innovations. While DMP may not have a direct connection with any Key Exploitable Result, all data outcomes should be aligned with Intellectual Property Rights (IPR) guidelines.

1.3 Deliverable Structure

The structure of this document is based on the template provided by the European Commission (EC), ensuring alignment with its guidelines and principles.

Section 2 describes the type of research data that will be gathered, processed, and/or generated within REALLOCATE, offering a comprehensive overview of their categories, formats, origins, frequency, and so forth. Section 3 stands as the core of this document. It presents how the project intends to make data Findable, Accessible, Interoperable, and Reusable, while harmonizing between the EC's requirements and the IPR strategy. Section 4 details the operational tasks related to Data Management, focusing on the workplan alignment and designating the partners responsible for overseeing the processes. Initially, the datasets will be presented with a broad overview, but as the project progresses, the details will become more specific. Section 5 presents the perspective of the Open Research Data Pilot (ORDP) and some indicative repository choices that can support the implementation of the pilot. Section 6 introduces a template for reporting the research data utilized throughout the project. Section 7 showcases the Comprehensive Knowledge Archive Network (CKAN) platform's capabilities for publishing, sharing, discovering, and utilising datasets. The methodologies that ensure data security during storage are described in

Section 8. Section 9 outlines the management of personal data generated within REALLOCATE, cross-referencing to corresponding deliverables. The Ethical considerations are described in Section 10 while Section 11 involves the next versions of the DMP indicating the dynamic nature of the document.

2   Data Summary

REALLOCATE will involve the collection and generation of diverse data categories to support its objectives:

Data that will be collected for analysis and assessment within REALLOCATE in order to accomplish its objectives include the following categories:

- **Mobility Data:** This category includes information on how people move within urban areas. It covers origin-destination data, travel behaviour surveys, traffic flow, public transport usage, active modes like cycling and walking, and vehicle-related data.
- **Environmental Data:** These data sources pertain to the urban environment's conditions. They include air quality, noise pollution levels, emissions data, climate and weather patterns, biodiversity assessments, and green space metrics.
- **Social Inclusion Data:** Data on social inclusion and demographics are vital for understanding the impact of mobility interventions on diverse population groups. This category comprises demographic and socioeconomic data, accessibility metrics, social surveys, and health and wellbeing indicators.
- **Spatial data:** This category involves spatial data, infrastructure layouts, street design, and 3D models of urban spaces.
- **Risk and Incident Metrics:** This category involves data related to urban safety, such as reported accidents, danger zones and security measures.
- **Governance and Policy Data:** Understanding the regulatory landscape and governance framework is crucial. This data category includes policy documents, regulatory information, and insights obtained from key stakeholder interviews and surveys.

The data will be available in diverse formats, including tabular data in CSV (Comma-Separated Values) and Excel formats, geospatial data in GIS (Geographical Information System) formats, satellite data (e.g. Copernicus database), sensor data from IoT (Internet Of Things) devices, textual data in the form of reports and documents, multimedia data including images and videos, and time-series data.

Also, the aforementioned data will have diverse origins. A significant portion of the data will be collected from the pilot cities participating in the project. These cities will provide a valuable insight related to their mobility infrastructure, incidents, accessibility and social inclusiveness. In addition, the project will leverage the results of surveys and questionnaires to capture information about daily activities, travel behaviours and the needs of the urban

population. Furthermore, existing public datasets will contribute to the project's knowledge base as they can be used as a preliminary dataset for the AI (Artificial Intelligence) algorithms. These datasets may include demographic information, traffic patterns and environmental factors. Also, environmental sensors and monitoring devices will play a vital role in collecting real-time information.

More specifically in the REALLOCATE project, data collection is structured around Key Performance Indicators (KPIs), each identified by a unique code to monitor specific aspects of the project's interventions. These KPIs are grouped into five thematic areas: **Safe & Sustainable Schools**, **Concepts for Space Reallocation**, **Data Safety Digital Integration for Accessibility**, **Central Areas Traffic Reorganization**, and **Integrated Traffic Reorganization – Peri-urban**.(and are described in more detail in the following sections) Participating cities selected KPIs relevant to their local context, specifying the data format, sources, and collection frequency that best suit their needs.

2.2 Safe & Sustainable Schools

This thematic area relates to school area improvements through engagement and enhancing public spaces around them. The KPIs under this category focus on community engagement, safety, infrastructure improvements, and environmental impacts. Cities such as **Utrecht**, **Bologna, Lyon** and **Warsaw** have been active in collecting data for these indicators.

- **Participation Rate**: This KPI measures the level of community engagement in developing and managing school environments. Participation is frequently tracked, capturing the number of people involved in relevant activities. Data collection methods vary, with some cities logging participation through event records, while others gather data during monthly workshops.

- **Incident Reduction**: This KPI tracks safety improvements by monitoring the reduction of accidents, near misses, and safety incidents around schools. It offers valuable insights into the quality of public spaces, particularly following safety audits and interventions. Data is sourced from police reports and traffic incident records, typically collected on a quarterly basis.

- **Bike Parking Capacity**: This indicator monitors the increase in the number of bike parking spaces or bike-sharing docks installed. It supports sustainable mobility by encouraging cycling as a safe mode of transport for students. The number of

additional bike parking spaces or docks is counted annually, using city infrastructure data and site surveys as data sources.

- **Stakeholder Feedback**: This widely used KPI gathers feedback and insights from stakeholders on the potential for extending and replicating interventions in other areas. Data is collected through surveys, interviews, and sentiment analysis, with semi-annual workshops held to engage schools and citizens. The real-time feedback helps keep the project aligned with community needs.

- **Road Space Reallocation**: This KPI measures the percentage and area of road space reallocated to improve active mobility and safety for pedestrians and cyclists. Data, typically collected annually, is drawn from spatial data, planning documents, and measurements of existing road space allocation.

- **Environmental Impact**: This KPI assesses the environmental improvements resulting from converting impermeable surfaces to permeable or vegetated areas. It focuses on enhanced stormwater management, the reduction of the urban heat island effect, and increased biodiversity. Metrics such as air quality and climate data are collected via sensors (measured in µg/m³) alongside environmental studies. The quality of the conversion is evaluated (e.g., High, Medium, Low), and data is collected monthly through city planning records and physical measurements, tracking changes before and after pilot interventions.

2.3 Concepts for Space Reallocation

This category focuses on optimizing the use of road space to create more cyclist- and pedestrian-friendly environments, with a priority on environmental sustainability. The KPIs within this thematic area are designed to measure improvements in infrastructure, address conflicts between road users, and promote green spaces. Participating cities in this category include **Barcelona, Budapest**, and **Heidelberg**.

- **Increased Bike Parking Spaces**: This KPI tracks the percentage increase in dedicated bike parking spaces in targeted areas. The goal is to enhance cycling and e-bike infrastructure by reducing waiting times and expanding available bike parking. Cities like Budapest and Heidelberg measure this KPI annually using site surveys and city infrastructure records, with the results reported as a percentage increase.

- **Addressing Conflicts Between Cyclists and Pedestrians**: This KPI monitors the reallocation of road space from motor vehicles to pedestrians and cyclists by

transforming streets into low-traffic or car-free zones. It measures the percentage of road space reallocated, and the number of incidents reported between cyclists and pedestrians. Cities collect data through police reports, surveys, spatial analysis, and safety audits. Data is recorded either annually or quarterly and reported through qualitative assessments.

- **Vegetated Surface Area Increase**: This indicator tracks the total area of newly added vegetated surfaces, including green spaces and planted areas. Cities measure this KPI annually in square meters (sqm) using spatial data and environmental surveys, drawing from city planning records.

- **Number of Trees Planted**: This metric records the number of trees planted in pilot areas, contributing to the greening of urban spaces. Data is collected annually through city planning records and environmental surveys.

- **Participation Rate**: As described under the **Safe & Sustainable Schools** category, this KPI measures community engagement in the project. The data collection methods and frequency are similar.

- **Stakeholder Feedback**: Also detailed under the **Safe & Sustainable Schools** category, this KPI collects feedback and insights from stakeholders to assess the potential for extending and replicating interventions.

- **Perceived Safety Improvement**: This KPI is designed to assess initial perceptions of safety in shared spaces following road reallocation measures. Cities gather data using user surveys and participatory workshops. Feedback from these sources is crucial for monitoring ongoing safety and making necessary adjustments.

- **Perceived Public Space Improvement**: This KPI measures the perceived enhancement in public spaces, as reported by local residents, school communities, and visitors, following road space reallocation and climate adaptation measures. Data is collected annually through surveys conducted with the local community and school participants.

- **Environmental Impact**: As previously discussed under the **Safe & Sustainable Schools** category, this KPI assesses the broader environmental benefits of the interventions, with data gathered through similar methodologies.

2.4 Data Safety Digital Integration for Accessibility

This thematic area focuses on the secure integration of traffic data and privacy measures, ensuring that data protection and accessibility are prioritized. The KPIs under this category are designed to measure stakeholder engagement, the use of services, and improvements in traffic safety. Key cities involved in this thematic area include **Barcelona, Lyon, and Tampere**.

- **Participation Rate**: As described under the **Safe & Sustainable Schools** category, this KPI tracks community engagement and participation in project activities, following similar data collection methods and frequencies.

- **Stakeholder Engagement Level**: This KPI evaluates the degree of engagement and participation from various stakeholders, including citizens, community groups, and organizations, in the simulation processes and feedback provision. It also assesses public acceptance of the interventions. The qualitative assessment (e.g., High, Medium, Low) is based on the extent to which stakeholders are involved and their input is incorporated into decision-making processes. Cities conduct semi-annual qualitative feedback collection and sentiment analysis to measure the initial level of stakeholder engagement.

- **Shared Journeys for Persons with Reduced Mobility (PMR)**: This KPI tracks the usage of shared journey services by persons with reduced mobility. In Barcelona, this KPI monitors the number of PMR users served, with data collected through monthly reports.

- **Community Participation Rate**: This KPI measures the percentage of local citizens and stakeholders who actively engage in the decision-making process or use the services implemented as part of road space reallocation and active mobility planning. Data is collected either annually or semi-annually, using participation records and community engagement logs to track the initial percentage of engaged citizens and stakeholders.

- **Safety Improvement**: This KPI assesses improvements in traffic safety and risk management, with data sourced from safety audits and accident reports. The KPI tracks traffic risk scoring and is reported quarterly.

REALLOCATE Deliverable 1.4 Data Management Plan

- **Incident Reduction**: As detailed under the **Safe & Sustainable Schools** category, this KPI monitors the reduction in accidents and safety incidents, using similar data sources and collection methods.

- **Integration of Different Data Sources**: This KPI focuses on the inclusion of multiple data sources to enhance the quality and accuracy of the collected data, particularly in relation to the MAIS3+ standard. It monitors the integration of data from systems such as environmental monitoring stations, with annual data quality reports used to assess progress.

2.5 Central Areas Traffic Reorganization

This thematic area focuses on reorganizing traffic in central urban areas to reduce accidents and enhance pedestrian safety through traffic calming measures and infrastructure redesigns. Key cities involved in this initiative include **Gothenburg and Zagreb**, both of which use Key Performance Indicators (KPIs) to monitor the effectiveness of pedestrian-friendly interventions and the reduction of traffic incidents.

- **Improved Walkable Conditions**: This KPI measures the percentage increase in the quality of walking conditions, such as sidewalk maintenance, accessibility, and pedestrian amenities, compared to baseline conditions before project implementation. It is assessed annually using walkability audits and infrastructure surveys, with results reported based on qualitative improvements identified through surveys and audit reports.

- **Adaptability Rating**: This KPI evaluates the potential for replicating a local solution in other urban environments. It is measured annually through expert assessments and feedback from stakeholders, with the aim of determining the scalability and adaptability of the interventions.

- **Walkability Index**: This KPI provides a score that reflects how accessible and pedestrian-friendly a redesigned urban area is, particularly for disabled individuals. The index is scaled from 0 to 100, and data is collected annually through walkability assessments and surveys conducted with local residents and users.

- **Accessibility Improvements**: This KPI monitors the progress in creating more inclusive and accessible public spaces for vulnerable road users (VRUs), such as pedestrians and cyclists. It tracks the percentage improvement in accessibility, measured annually through accessibility surveys and VRU counts.

- **Participation Rate**: As described under the **Safe & Sustainable Schools** category, this KPI measures the level of community engagement in project activities. Data collection methods and frequency are consistent with those described earlier.

- **Stakeholder Feedback**: Similarly, this KPI is used to gather insights and feedback from stakeholders regarding the interventions. See the **Safe & Sustainable Schools** section for details on data collection and reporting.

- **Safety Improvement**: This KPI measures enhancements in traffic safety following interventions. See the **Safe & Sustainable Schools** category for further details on the data sources and methods used to track safety improvements.

- **Pedestrian and Cyclist Comfort Index**: Refer to the **Walkability Index** for further information on how this indicator is measured and reported.

2.6 Integrated Traffic Reorganization – Peri-Urban

This thematic area addresses the management of traffic in peri-urban regions, with a particular focus on the integration of public transport systems and improving overall traffic flow. Key cities participating in this category, including **Budapest, Gothenburg, and Heidelberg**, are working to enhance public transport usage and reduce dependence on cars. The following Key Performance Indicators (KPIs) have been established to monitor the effectiveness of these interventions:

- **Cyclist Count**: This KPI tracks the total number of cyclists recorded in the project area, either daily or hourly, using methods such as traffic cameras, sensors, or manual counting. The data provides insights into cycling activity levels and is recorded continuously or at set intervals throughout the year.

- **Active Mobility Mode Share**: This KPI measures the proportion of trips made by pedestrians and cyclists compared to the total number of trips in the project area. It reflects the shift towards active mobility modes and is recorded annually using traffic surveys and observational data from transportation monitoring systems.

- **Community Engagement Index**: This KPI assesses the level of community acceptance and satisfaction with the safety measures implemented as part of the traffic reorganization. The index is derived from annual surveys, interviews, and community engagement logs, providing a qualitative assessment of community response to the interventions.

- **Mode Share Shift**: Similar to the **Active Mobility Mode Share**, this KPI quantifies the shift away from motorized transport modes (e.g., cars, motorcycles) towards active and public transport modes. It is measured annually using traffic surveys and other observational data, with a focus on sustainable transportation practices.

- **VRUs' Perception of Safety**: This KPI captures the perceived safety of vulnerable road users (VRUs), including pedestrians and cyclists, following the traffic reorganization. The sentiment is gathered annually through surveys and interviews, with a focus on understanding how safe VRUs feel in the redesigned traffic systems.

- **Perception of Impact**: This KPI evaluates the acceptance and sentiment of citizens and stakeholders regarding the urban design interventions implemented in their community. The data is collected through annual surveys and interviews, offering insights into public perception of the project's impact.

- **Accessibility Improvement**: Refer to the **Accessibility Improvements** section under **Central Areas Traffic Reorganization** for details.

- **Green Area Expansion**: This KPI tracks the expansion of green spaces in the project area, measured in square meters. The data is collected annually using satellite imagery and land use surveys, providing a quantitative assessment of newly added vegetated areas.

- **Feasibility for Replication**: Refer to the **Adaptability Rating** section under **Central Areas Traffic Reorganization** for information on how this KPI is measured.

- **Incident Reduction**: Refer to the **Incident Reduction** KPI under **Safe & Sustainable Schools** for details on data collection and analysis.

- **Environmental Impact**: For details on the environmental impact assessment, refer to the **Environmental Impact** KPI under **Safe & Sustainable Schools**.

An overview of the categorisation of the KPIs into the five thematic clusters is also illustrated in Figure 1.

2.7 Other Generated Data

Data generated and illustrated in the project deliverables represent a significant category of data produced during the project. These data will be available through the project's official page https://reallocatemobility.eu. The publication of deliverables will occur
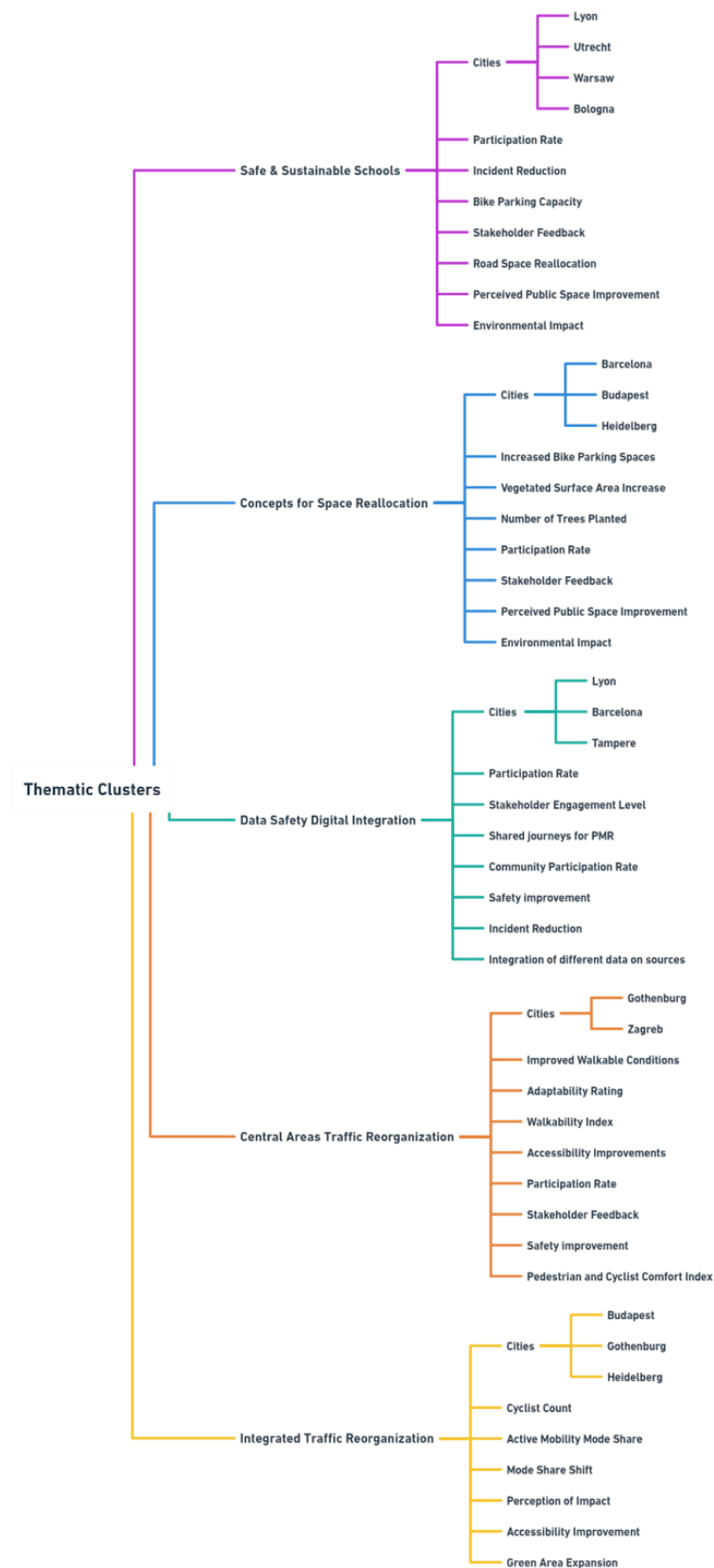
Figure 1 Common Indicators per Thematic Cluster (TC) and the included SSMLs. (Source: Deliverable 5.1-Evaluation & Impact framework of the project)

REALLOCATE Deliverable 1.4 Data Management Plan

upon formal approval by the project consortium and relevant authorities and only public deliverables will be published online. These documents will be provided in PDF (Portable Document Format) with enabled markers for straightforward navigation and retrieval.

Publications in scientific journals, conferences, articles in magazines, project leaflets and posters will play a pivotal role in disseminating REALLOCATE's research outcomes and technical advancements. Through these activities, REALLOCATE aims to showcase its methodologies, findings and solutions, contributing to the collective knowledge base and fostering collaboration with experts and practitioners in the field. For certain categories of outputs like participation in exhibitions, brochure guides etc. the project will utilize its social media platforms to make announcements.

The developed software and front-end application suites which will be the outcomes of the project and will be presented in the deliverables and be visualized through the REALLOCATE AI-powered Distributed Dashboard. Also, the IPR and potential for exploitation of the source code will be examined as the project progresses.

The information provided in this Section represents a preliminary overview of the DMP for the REALLOCATE. It's important to note that the DMP will undergo further refinement and reach its final version in M36 of the project. At that stage, more detailed information will be provided regarding the characteristics of each dataset and the handling of IPR.

3 REALLOCATE FAIR Management Principles

To ensure easy accessibility, downstream use, and the reusability of research data derived from REALLOCATE, FAIR Data Management principles and guidelines are adopted. These principles are essential for enhancing data reusability and knowledge integration, enabling other researchers and individuals to benefit from and expand upon the research while improving their own results.

3.2 Data Findability and Provision for Metadata

In line with the commitments outlined in the Grant Agreement (GA) and Consortium Agreement(CA), which define the management of results, access rights, and information non-disclosure, trusted open-access data repositories such as CKAN or other compatible tools (e.g., Zenodo, IEEE DataPort, GitHub) will be utilised to ensure data findability. For metadata management, specific requirements to ensure adherence to FAIR principles should be followed. The metadata must provide detailed information, including dataset descriptions, deposit dates, author details, funding acknowledgments, the project name and number, licensing terms, and persistent identifiers (PIDs). This ensures that all research outputs are accessible, traceable, and reusable, promoting transparency and long-term data usability.

Persistent Identifiers (e.g., Digital Object Identifier or DOI) will be employed to access scientific publications (papers). DOIs, as globally unique character strings consisting of numbers, letters, and symbols, serve to permanently identify electronic documents and provide web links through URLs. Metadata associated with the DOI name is stored and can be cited under the same URL. The DOI remains unchanged even if the document's URL changes, ensuring stable linkage. REALLOCATE's papers will have DOIs assigned by their publishers or repositories, while others may be assigned by the entities managing the data repository. In some cases, the REALLOCATE Consortium may need to obtain the identifier, and services from a DOI Registration Agency (DRA) will be employed for this purpose. DRAs collect metadata, assign DOI names, and provide services such as reference linking and metadata lookup.

Each public REALLOCATE dataset will be linked to a unique URL identifier associated with the corresponding DOI. These URLs will include the actual download link for the dataset,

ensuring accessibility from various websites. Furthermore, all documents posted on the project's website will be securely archived in REALLOCATE's private archive.

Clear guidelines have been established for naming and versioning REALLOCATE documents. Table 1 illustrates the naming strategy and versioning for REALLOCATE documents.

Table 1. Naming strategy and versioning for REALLOCATE documents.

Document	Naming Strategy
Deliverables REALLOCATE _Deliverable_ – [Name]	REALLOCATE_Del_Name
Example	REALLOCATE_D1.1
Internal Del. Evaluation	REALLOCATE _Del_Evaluation_ – [iterative evaluation number]
Example	D1.1_Evaluation_2
Financial Reporting	REALLOCATE _Financial reporting_Mx-My_PXX – [partner acronym]
Example	REALLOCATE _Financial reporting_M1-M18_P01-CERTH
Technical Reporting	REALLOCATE _Technical reporting_Mx-My_PXX_[Partner Acronym]
Example	REALLOCATE _Technical reporting_M3-M4_P01-CERTH
Agenda of Meetings	REALLOCATE _Agenda_YYYYMMDD
Example	REALLOCATE _Agenda_20231031
Meetings Minutes	REALLOCATE _Minutes_YYYYMMDD
Example	REALLOCATE _Minutes_20231031

REALLOCATE Deliverable 1.4 Data Management Plan

Also, the type of the meeting can be defined as follows.

- **Type of meeting:** Acronym for the type of meetings (both physical and by other means) that can be conducted in the project:
  - Plenary
  - AB – Advisory Board meeting
  - WP#No – Work Package meetings
  - TX.Y – Task meetings,
  - etc.

Changes in the documents should be registered in the History table along with other information relevant to the change (date, version number and the kind of modification made).

3.3 Data Accessibility

Upon approval of project deliverables for public release by the REALLOCATE Consortium, the European Commission, and external reviewers, they will be made available for download on the REALLOCATE project website (https://reallocatemobility.eu/resources). This will enhance access and facilitate the dissemination of project knowledge. These deliverables will be uploaded to the website in PDF format immediately following their acceptance by the European Commission.

For deliverables with high levels of confidentiality, they will not be publicly available on the website but will be accessible only through internal communication channels within the Consortium. Special attention should be given to this category of deliverables to ensure they are protected during any embargo period due to IPR, such as Know-How or patents, or for verification purposes, along with the provision of software documentation and reference materials required to interpret them.

For scientific publications and in alignment with Horizon Europe's open-access requirements, all peer-reviewed publications funded through this project will be made openly accessible. If, for critical reasons, certain publications need to adhere to restrictive rules imposed by Consortium members, publishers of scientific papers, or face copyright limitations from editorial companies, the strategies of 'Green Open Access' or 'Gold Open Access' may be adopted. Gold Open Access implies publication on an online open-access publication site, while with green, the paper is initially published in a written journal and then self-archived online. Furthermore, pre-print versions of papers, prior to peer review, can be

released in the form of Technical Reports, which will be published on the project website or on institutional and preprint servers (e.g., Zenodo, arXiv, F1000 Research). More detailed information regarding the scientific publication processes will be specified in alignment with the dissemination plan.

As a general guideline, REALLOCATE is committed to adhering to the FAIR data management principles for any document disseminated through online data repositories. The appropriate means of publication for each individual work will be assessed and decided upon accordingly.

Document workspace and documentation:

The workspace-application selected for the project's everyday-work include Google Drive managed under UCD control and Eurocities SharePoint. Google Drive provides syncing and sharing of files of different types. It refers to a suite of client-server software for creating and using file hosting services. It can be utilised free of charge by offering several functionalities that are needed to support the execution of the project, including document editing, content management, calendar and task organization, contact lists and document management - accessible to users individually authorized by UCD.

Furthermore, UCD hosts project files on Google Drive using Google Workspace for Education Plus, a secure and compliant solution tailored for data storage and management. UCD's contractual agreement with Google ensures the platform meets strict standards for security, data protection, and ongoing monitoring. This transition in 2019 focused on implementing robust data governance and enhanced security protocols. With Google's network of EU-based data centers, UCD's data remains within European borders, ensuring compliance and reliable data accessibility for all project partners.

In addition, Eurocities SharePoint is used as a collaborative workspace that facilitates file sharing, content management, and communication among partners. It supports document versioning, co-authoring, secure access control, and centralized storage for project-related documents, ensuring easy access and collaboration across the consortium. SharePoint also provides additional features such as team sites, task management, and integration with Microsoft Office tools, further enhancing the project's document management and communication capabilities.

Software repository and documentation: REALLOCATE's Multimodal Data Hub and Distributed Dashboard will act as central repositories for software resources and documentation in accordance with the grant agreement's provisions. These platforms will serve as accessible entry points for users, offering tailored processed data, research

findings, and an array of tools and services developed and tested within the SSMLs. Moreover, the adopted strategy guarantees GDPR compliance for data collected centrally at the Data Hub, allowing data sharing within the consortium and with the NetZeroCities Platform, ensuring the long-term maintenance and accessibility of acquired data. Non-sensitive processed data will be securely stored with robust security measures, including encrypted data exchange and public key authentication, alongside redundancy protocols for storage. Non-sensitive raw data will be removed from servers after the project's conclusion, unless needed for future analysis, in which case it will be archived in Zenodo. Additionally, data and research outputs will remain openly accessible for a minimum of five years.

SRC repositories as GitLab/GitHub: REALLOCATE employs SRC repositories like GitLab and GitHub to create a secure and collaborative environment for code management, seamless sharing of local file changes, and code exchange. Utilizing these SRC repositories guarantees that the information remains easily accessible, usable, and open to contributions among the IT partners, thereby promoting collaboration and fostering innovation within the project.

Docker, Binder: Open-source software platform enabling creation, deployment and management of virtualized application containers on a common operating system independent from the host machine.

Software assets open to public: REALLOCATE is committed to providing accessible access to its software resources and research outcomes. Following the principles outlined in the 'FOSTER open science taxonomy' and utilizing appropriate licensing options like Creative Commons, Open Data Commons, and others, the project aims to make its results available to the wider community, including project partners. Virtual research environments like the European Open Science Cloud, Open Science Framework, and collaborative writing platforms such as Overleaf and Authorea, along with professional and academic social networks like ResearchGate, Academia.edu, Loop, and LinkedIn, are some of the platforms that can be used to disseminate these resources to researchers worldwide. This strategy ensures that REALLOCATE's software assets and research findings reach a broad and geographically diverse audience, promoting collaboration and knowledge sharing among researchers from various regions. Certain software outcomes, including source code and AI/ML models, may have restrictions on their dissemination to the open public within REALLOCATE, primarily due to IPR limitations or the need to protect potential exploitation by project partners. Throughout the project's implementation, the legal and contractual framework will establish specific boundaries regarding the level of access and utilization of project outcomes.

3.4 Data Interoperability

In this Section are presented the recommended types of data formats in order to facilitate the interoperability of data. The project will consider employing common standards such as jpg and mp4 for images and videos, while point clouds will be stored in ASCII file formats like XYZ, OBJ (with some exceptions), PTX (Leica), and ASC. For other data types, CSV files will be utilized. Additionally, simulation and trial parameters, as well as algorithmic codes, will be provided in txt files. As part of our data management strategy, the open-source CKAN [1] platform will be used ensuring seamless integration and accessibility of various data types. A detailed breakdown of the recommended and acceptable formats for each data type can be found in Table 2.

Table 2. Recommended data formats to enhance interoperability.

Textual documents
Recommended formats: Plain text, ASCII (.txt) XML (.xml) JSON (.json) Adobe PDF (.pdf) **Acceptable formats:** Hypertext Mark-up Language (.html) MS Word (.doc/.docx) Software-specific formats (.odt, .ppt, .yaml)

Databases
Recommended formats : Comma-separated values (.csv) MS Excel (.xls, .xlsx) Clear text files (.txt) Machine formats (.json) **Acceptable formats:** SQL (.sql) Other formats associated to specific database software

Image
Recommended formats: JPEG (.jpeg, .jpg) GIF (.gif) PNG (.png)

REALLOCATE Deliverable 1.4 Data Management Plan

Acceptable formats: TIFF (.tif, .tiff) Photoshop files (.psd) BMP (.bmp)

Audio

Recommended formats : FLAC (.flac), WAV (.wav)

Acceptable formats: Audio Interchange File (.aif)

Video

Recommended formats : MPEG-1 Audio Layer 3 (.mp3), MPEG-4 (.mp4) OGG video (.ogv, .ogg) Motion JPEG 2000 (.mj2)

Acceptable formats: WAV (.wav), AVCHD video (.avchd)

Software

Recommended formats : Docker images reference

JAVA .jar, .war

Scripts .sh, .bash, .py

Project format following usual practices (https://guides.github.com/features/wikis)

Acceptable formats: Other extensions of source code

Point Clouds

Recommended formats: ASCII file formats (.xyz, .obj with some exceptions), PTX (Leica), ASC

Acceptable formats: Other formats associated with point cloud data

3.5 Data Re-usability

Reusability of the project outputs can be acquired and be source for other projects through deliverables, papers, presentations and project results published on the project's website. Moreover, use of common types of licenses for data sharing/re-use (e.g. Creative Commons, Open Data Commons), use of open tools for data interpretation/re-use (e.g. Open ASCII editors (e.g. Notepad++), Open image editing and visualization tools (e.g. GIMP, etc.), provision of documentation for data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, etc.) and adoption of relevant data quality assurance processes to be applied.

An indirect access to some of project results could also be possible via specific standard and technical organizations (Standard Developing Organizations –SDOs). However, the rules applied to each one vary. Either they allow open access to general public (ETSI) or limited access only to members (IEEE). Additionally, some results can be accessed but cannot be reproduced / used without permission. Specific confidential material will require direct licensing from the originating company.

Data produced within REALLOCATE that will be openly published on the website, will be useable by external stakeholders, during and after the end of the project. In case of use of REALLOCATE data, there is a requirement for appropriate attribution back to the project. In case of modifications to the original data or results, this should be stated clearly. Data will remain accessible for as long as the project website is kept open (3 years after its ending).

Restrictions for the reusability of the project outcomes will be updated to the final version (D1.5) of the deliverable if required.

4 Allocation of Resources

The execution of the Data Management Plan in REALLOCATE will be carried out as part of task T1.6 - Data management, with a particular focus on delivering D1.3 - Data management plan in its initial version which has already been submitted. Subsequently, updates to the plan will be provided in the current deliverable D1.4 (at month 18) and D1.5 (at month 36) in their respective updated and final versions.

CERTH will hold the responsibility for the DMP-Task 1.6, which will include comprehensive details covering various aspects. These aspects include data collection, curation, storage, long-term preservation, security, quality assurance, allocation of Persistent Identifiers (PIDs), provision of metadata aligned with disciplinary requirements, licensing, and rules and procedures for data sharing. The project's data management philosophy revolves around adhering to FAIR data management principles, ensuring that data is readily accessible for various purposes, such as supporting data-driven decisions and investigating AI predictive models. CERTH will take charge of guiding and implementing critical data processing steps, focusing on data originating from 10 SSMLs. These steps will be conducted in alignment with the FAIR guiding principle and will adhere to the European GDPR and related national legislations.

In order to execute the tasks outlined above, a thorough budget assessment has been conducted. The estimated total resources allocated to this endeavor have been carefully calculated, drawing from contributions from various Consortium partners. This allocation ensures a substantial commitment of personnel and resources towards effective data management, encompassing aspects such as protection, security, documentation, licensing, and information reuse. All these efforts remain in strict compliance with the regulations and principles governing FAIR data management.

The long-term preservation of each dataset will be a subject of discussion during the implementation of Task T1.6, with further analysis and planning taking place as the project progresses.

5 Open Research Data Pilot (ORDP)

The Open Research Data Pilot of the European Commission enables open access and reuse of research data generated by REALLOCATE project. The EC recognizes that there are good reasons to keep some or even all research data generated in a project closed. That is why, ORDP is "invented"; to balance openness and protection of scientific information, commercialisation, Intellectual Property Rights, privacy concerns and security. The definition of ORDP refers to a flexible pilot in order to improve and maximise access to and re-use of research data generated by the project.

The ORD pilot applies primarily to the data needed to validate the results presented in scientific publications. The beneficiaries can provide also other data voluntarily.

"Participating in the ORD Pilot does not necessarily mean opening up all research data. The principle "as open as possible, as closed as necessary" and focuses on encouraging sound data management as an essential part of research best practice" [2].

5.2 ORDP Datasets

Open source datasets and data extracted from the analysis within the project are planned to be openly accessed to enable distributing of the results to the community and that they can be freely accessed, mined, reproduced and disseminated by third parties. All partners have the possibility to opt out of this process, depending on IPR issues, personal exploitation plans and data ownership.

Concerning the publications category, all of them will be made open access type gold (immediately accessible for free) if not type green. In the latter case, they will be immediately released after the embargo period. It should be noted here that if a peer-reviewed publication contains any commercially sensitive content it will pass through IPR screening before being published and if any publishers are not "open access friendly", REALLOCATE can always opt to publish pre-print forms of articles as open access when allowed by the publishing companies. However, it is the commitment of REALLOCATE to publish any scientific outcome following the "green open access" approach.

At this phase of the project, the above references to the way the ORDP will function are preliminary and may change as the projection evolves.

5.3 Repositories

All data intended for sharing will be stored and managed using the CKAN platform, an open-source data management system selected for its robust capabilities in data accessibility, metadata support, and compatibility with best practices for data publication and sharing (see Section 7 for details). CKAN will enable project data to be accessed, mined, and reused by third parties and is designed to meet legal and security requirements for data management, supporting long-term storage and discoverability.

Additional repositories, such as the REALLOCATE website – Publications tab (https://reallocatemobility.eu/), may also be used to ensure broad accessibility to publications and other research outputs as the project evolves. For internal sharing among REALLOCATE partners, datasets will be accessible through the project's document repository on Google Drive, managed by the Project Coordinator.

6 Data Management Monitoring Template

The datasets that will contribute to the project's research and evolution should be recorded, so as to be in alignment with FAIR management principles as declared in the former sections of the DMP.

For the configuration of the various datasets that will be used in the project, a template is set where the partners should report the datasets that will help their research.

Figure 2 illustrates the template for the dataset reporting. The parameters of the datasets that will appear in the project should be described, giving initially general information and in greater detail as moving towards the end of the project. It aims at representing all the parameters of a datasets useful for data management, including size, reusability, metadata, size, etc.

Reference/name			Type of data sharing		Public/Private/etc.
Description			To whom it could be useful		Researchers, scientific community on the field
Type			Related/Indicative similar datasets		
Related WPs and tasks			Possible synergies with similar data		
Format, standards			Releated publications (if any)		
Necessary S/W and other tools for enabling re-use					
Estimated size (MB/GB/TB)			Back-up (method, frequency…)		
Storage (where)			Means for personal data protection		
Personal data included	Yes/No		IPR Owner/Data Owner		Partner PXX YYYY
Open Research Data Pilot (ORDP)	Yes/No			Yes/No	Yes/No
Link where it can be accessed (if exists)			Metadata	Adapted Format	eg. OpenAIRE

Figure 2 Data management monitoring template

Indicatively, the following table has been completed for the KPI related to Mobility Space Usage for the Barcelona Pilot 1 baseline action, using the proxy indicator "Conflicts between Cyclists and Pedestrians", as outlined in D5.1 – Evaluation and Impact Framework.

Table 3. Data management monitoring template for "Conflicts between Cyclists and Pedestrians" indicator for Barcelona Pilot 1 – Baseline Action.

Reference/name	Vehicles involved in accidents handled by the police in the city of Barcelona	Type of data sharing	Public
Description	List of vehicles involved in accidents handled by the police in the city of Barcelona . An accident	To whom it could be useful	Researchers, project stakeholders, city decision makers

	may be associated with more than one vehicle. Including whether the cause is pedestrian , vehicle type , model , brand, color and license type and age of the person who was driving.		
Type	Police report	**Relative/Indicative similar datasets**	Accidents managed by the Guardia Urbana in the city of Barcelona
Related WPs and tasks	WP5 (T5.1, T5.2) WP3 (T3.2)	**Possible synergies with similar data**	Extended analysis of accident data
Format, standards	CSV	**Related publications (if any)**	None
Estimated size (MB/GB/TB)	3.9 MB	**Means for personal data protection**	Pseudonymization
Personal data included	No	**IPR Owner/Data Owner**	Open Data BCN
Link where it can be accessed (if exists)	https://reallocate-ckan.iti.gr/dataset/vehicles-involved-in-accidents-handled-by-the-police-in-the-city-of-barcelona	**Metadata**	**Yes/No** Yes
			Adapted Format DCAT

7 Comprehensive Knowledge Archive (CKAN)

CKAN [3], which stands for Comprehensive Knowledge Archive Network, is a powerful, robust and adaptable open-source data management platform, specifically designed to meet the varied needs of organizations and individuals focused on publishing, sharing, discovering, and utilising datasets. CKAN emerges as an ideal platform, offering a comprehensive infrastructure, that streamlines and enhances data management processes, making it an ideal platform for organizations seeking scalable and efficient solutions to manage their data. Within the scope of REALLOCATE, CKAN has been implemented for the needs of T5.2 as the dashboard's database; additionally, its capabilities are utilised in DMP to optimize data management workflows, ensuring seamless and secure data handling while also boosting collaboration.

As the data platform is a critical element within the REALLOCATE project, functioning as a centralized data repository for all involved partners to securely store and share relevant datasets, a comprehensive literature review of *state-of-the-art* data repositories was conducted to identify the most suitable solution. The review considered several platforms, including CKAN, DKAN [4] (an open-source platform built on Drupal's Content Management System), Socrata [5] (a cloud-based enterprise software aimed at enhancing access to public data), and Junar [6] (a cloud-based data platform designed for organizations seeking efficient data publishing and sharing options) and evaluated them in terms of Security, Open-Source capability, API access, Data governance, Multi-modal data support and Extensibility. Table 4 presents an overview of the key features, along with an evaluation of how effectively each platform incorporates them. CKAN emerges as the most qualified platform for the purposes of the REALLOCATE project. A detailed comparison of all the aforementioned platforms and their functionalities is available in deliverable D5.2 – Distributed Dashboard & Centralised visualisation [7].

CKAN is widely recognized and utilised by governments [8], research institutions [9], and data-driven organizations globally [10], highlighting its robust, reliable and adaptable nature. Furthermore, CKAN is backed by a very active community of developers, dedicated to its continuous development and improvement. Specifically, the community is responsible to keeping CKAN resilient against security vulnerabilities, ensuring that it stays up-to-date and protected from both current and future threats. Focused on security, data integrity and user

needs, CKAN stands as an invaluable tool for any organization looking to effectively manage and share data securely.

Table 4. Comparison of Data Management Platforms

Feature/Aspect	CKAN	DKAN	Socrata	Junar
Security/Reliability	Strong	Good	Strong (Enterprise)	Basic
Open-source	Yes	Yes	No	No
API Access	Strong	Good	Strong	Good
Data Governance	Strong	Good	Good	Basic
Multi-modal Data Support	Yes	Yes	Yes (limited)	Limited
Extensibility	Strong	Good	Limited	Limited

To facilitate an easy and efficient use of all the functionalities of the CKAN platform utilized in the REALLOCATE project, a comprehensive user manual has been created as a guide for all participating partners. This manual is included in the Appendix II of the present file. It is also noteworthy that a description of the CKAN platform was provided in the Deliverable D5.2, where it was discussed as part of the AI-distributed dashboard's architecture, specifically as the primary source for dataset management and retrieval. In this deliverable, the CKAN platform is analysed with a focus on its role as a data repository for REALLOCATE, with particular emphasis on its key dataset-related functionalities.

7.2 Community

CKAN has a dynamic global community of developers, data enthusiasts, and institutional users who actively and collectively contribute to the platform's continuous improvement. This thriving community enhance the platform by developing new extensions, providing technical support and collaborating to address any feature/security gaps. Specifically, for new users, CKAN offers an extensive and detailed documentation, complete with step-by-step tutorials

that guide the users through the full range of the platform's functionalities. This ensures that even unexperienced users can easily navigate and harness all capabilities of CKAN, making it accessible to a broad audience.

A great example of the benefits of CKAN's active community is the flexibility it offers in terms of installation options. Users have the ability to install instances of CKAN for quick and easy development, or for more advanced users, they can perform source installation, which is ideal and recommended for development purposes and users with a deeper technical background. This versatility reflects CKAN's commitment of addressing access to both experienced and unexperienced developers, thus making the platform available to anyone.

The main advantage of CKAN is the platform's multi-modality and adaptability, with its extensible and modular architecture that can be fine-tuned and customized to meet the unique needs of different users. With access to a repository of over 200 community-developed extensions, CKAN users can easily tailor the data portal to suit various organizational and project-specific requirements. Moreover, CKAN provides an extending guide [11] for advanced users, enabling them to create their own, custom made, extensions that address specific needs, ensuring that the platform remains highly flexible and adaptable across its entire user base.

7.3 CKAN Platform Landing Page

A dedicated CKAN page has been developed to serve as the designated data repository for the needs of REALLOCATE project, and can be seen in Figure 3. This customized data repository provides seamless access to all of the CKAN's core functionalities and subsequent features, including publishing, sharing, discovering, and utilising datasets.

Through the repository's landing page, users can effortlessly navigate and take full advantage of CKAN's advanced data management capabilities. To its core, the repository is designed to streamline the data handling processes, providing a way to access all available data in an easy and organized way. By having a centralized data management platform, data transparency and usability is inherited to all authorized partners involved in the REALLOCATE project. These properties can be met by ensuring that all datasets are properly cataloged, easily searchable and available for reuse at any given time. Additionally, the platform's adaptability allows continuous growth and refinement, making it ideal to meet the evolving project requirements.
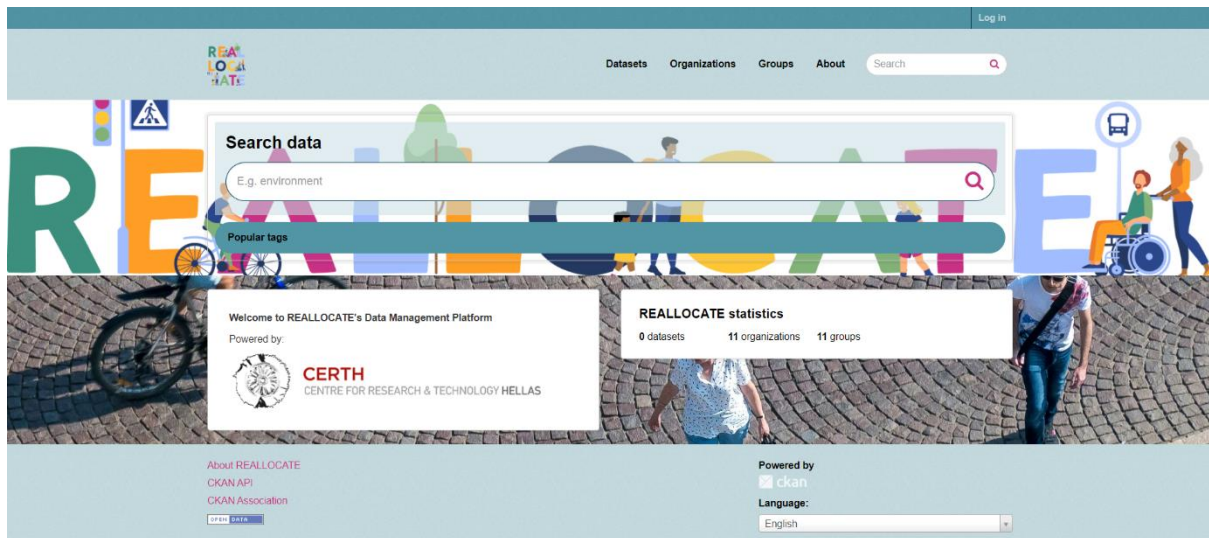
Figure 3 Landing page of the REALLOCATE data repository

Specifically, the platform offers a range of functionalities designed to ensure that users have a smooth and efficient navigation experience. These features include:

- **Datasets**: A dedicated page to datasets, where users can explore, publish and share data. Further details provided in 7.4.

- **Organizations**: A page dedicated to organizations, which are essential entities responsible for the ownership, management and distribution of datasets. Further details provided in 7.5.

- **Groups**: A page dedicated to groups, where related datasets are organized together, simplifying data access and enhancing the efficiency of data retrieval. Further details provided in 7.6.

- **Search bar**: A powerful tool that allows users to quickly and efficiently locate datasets, organizations, or groups by entering relevant keywords. This functionality is present in most aspects of the platform, so users can effectively find the information they need, no matter where they are within the system.

- **Statistics panel**: A real-time snapshot of the platform's current content, providing users with key metrics, such as the total number of datasets, organizations and groups available, offering this way a clear overview of the available resources.

The combination of all these functionalities is designed to significantly improve user experience, enabling seamless navigation and easy data discovery. In addition, the platform adheres to industry standard security protocols, ensuring that all data and user activities are safeguarded against potential threats.

7.4 Datasets

Datasets serve as the fundamental unit of the REALLOCATE data platform, since these are the building blocks that all involved partners will contribute and manage inside the platform. All datasets are accessible and can be managed through the datasets page, as seen in the Figure 4, where authorised users can view and interact with the full repository of available data. Detailed information regarding how the data platform complies with GDPR policies, data protection, secure storage and curation can be seen in Section 8.
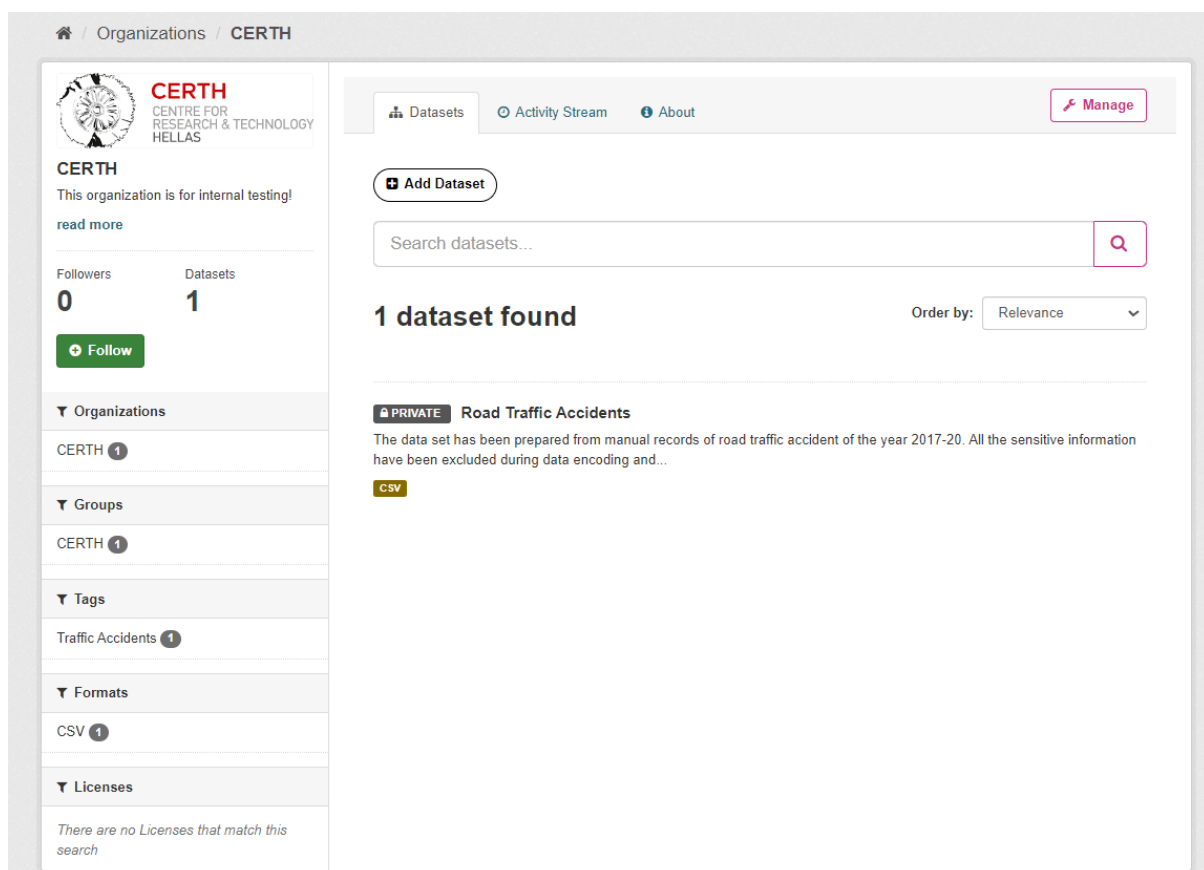
Figure 4 Datasets page of the REALLOCATE data platform

One of CKAN's most distinctive features is its well-polished search capability, which empowers users to quickly refine data searches based on specific criteria. Whether it's filtering datasets by data type, organization, or data format, users can swiftly pinpoint the exact datasets they need. This functionality can be accessed through the filtering options as seen in the left sidebar of Figure 4, and can effectively be used to narrow down vast amounts of datasets retrieved from the search result.

Moreover, CKAN also prioritizes data transparency and usability by allowing users to preview datasets through various visualization tools. Specifically, users can preview datasets

though interactive maps, graphs, and tables before deciding to downloading or integrating the data into their workflows. These previews provide an instance snapshot of the dataset's structure and content, and is especially beneficial for users looking to explore data trends/patterns to make informed decisions, without the need of external tools. An example of a grid preview can be seen in the left side of Figure 5, while an example of a graph preview is visible in the right side of Figure 5.
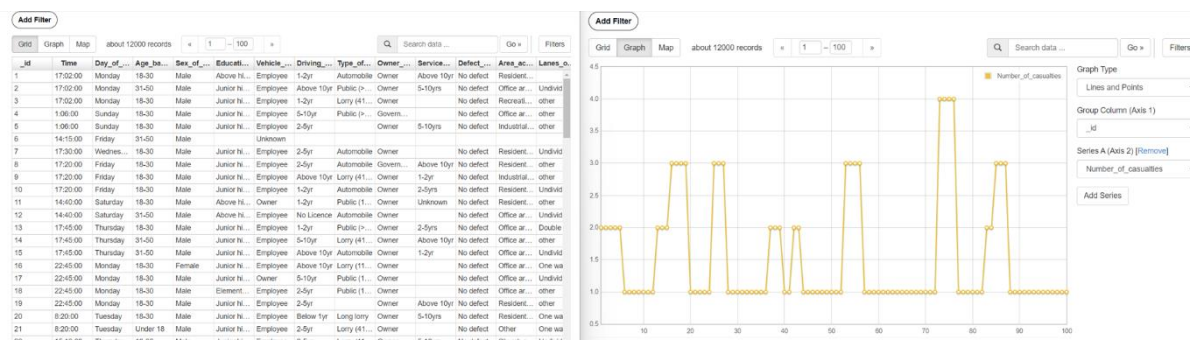
Figure 5 Preview of datasets within the REALLOCATE data platform

A very compelling feature of the platform, is CKAN's metadata-rich environment, since every dataset published is also accompanied by detailed metadata. These metadata include multiple information, some of which are the dataset's title, description, tags, licensing information, associated organizations, visibility status, data version, author and maintainer information. The existence of rich metadata provides users with the necessary means to fully understand the relevance, reliability and origin of the dataset they are accessing. This aspect of the platform is also very important for the maintenance of the data integrity and traceability of activities regarding uploaded datasets.

CKAN also supports multimodal data, which means that datasets of various formats (CSV, PDF etc.) can be hosted in the platform. This flexibility ensures that the platform can handle a wide range of data needs, enabling all involved partners to upload and share datasets that best suits their particular case.

It is important to mention that data access is controlled, with CKAN offering specific visibility settings to ensure data privacy and security. Dataset access is granted only to authorized individuals, thus, only the rightful owner of the dataset can specify who has the ability to access the respective data source. During the dataset upload process, users can select whether the dataset should be public (visible to anyone), or private (only visible to members of the same organization) by using the visibility field, ensuring that sensitive information is appropriately protected.

REALLOCATE Deliverable 1.4 Data Management Plan

7.5 Organizations

CKAN gives top priority on data security and privacy, ensuring that sensitive information is protected throughout the data management process. One of the main components of the platform is the use of organizations as the central mechanism of managing access control over the distributed datasets. Organizations in CKAN serve as gatekeepers, allowing users to share and distribute their dataset only to authorized personnel, thus, it provides well defined solutions for users to ensure data ownership and provide permissions where needed.

When users go through the dataset upload process, they are required to assign the dataset to the respective organization which owns it. As a result, all datasets are properly categorized and governed by the access policies of their particular organization. This assignment is particularly crucial for sensitive or confidential data, as it can limit the access exclusively to members of the organization, safeguarding this way the integrity of the data.

A visual representation of the organizations page can be seen in Figure 6, where users can browse and select any of the existing organizations to view additional information. While all users can view the list of all available organizations, data restrictions may still apply within the organization. Public datasets are openly accessible to everyone that has access to the platform, while private datasets are completely hidden from unauthorized individuals. As an additional safety measure, unauthorized individuals will not even be aware of the existence of any private dataset unless access is specifically granted. In case a user wants to access a private dataset of a specific organization, the organization administrator, or the site administrator should be contacted to process the request and grant the respective access where applicable.

To further enhance the protection and management of the existing datasets, roles are also assigned to individuals within the organization. These roles define the level of access and privileges that users have, ensuring that datasets are viewed and handled by trusted individuals.
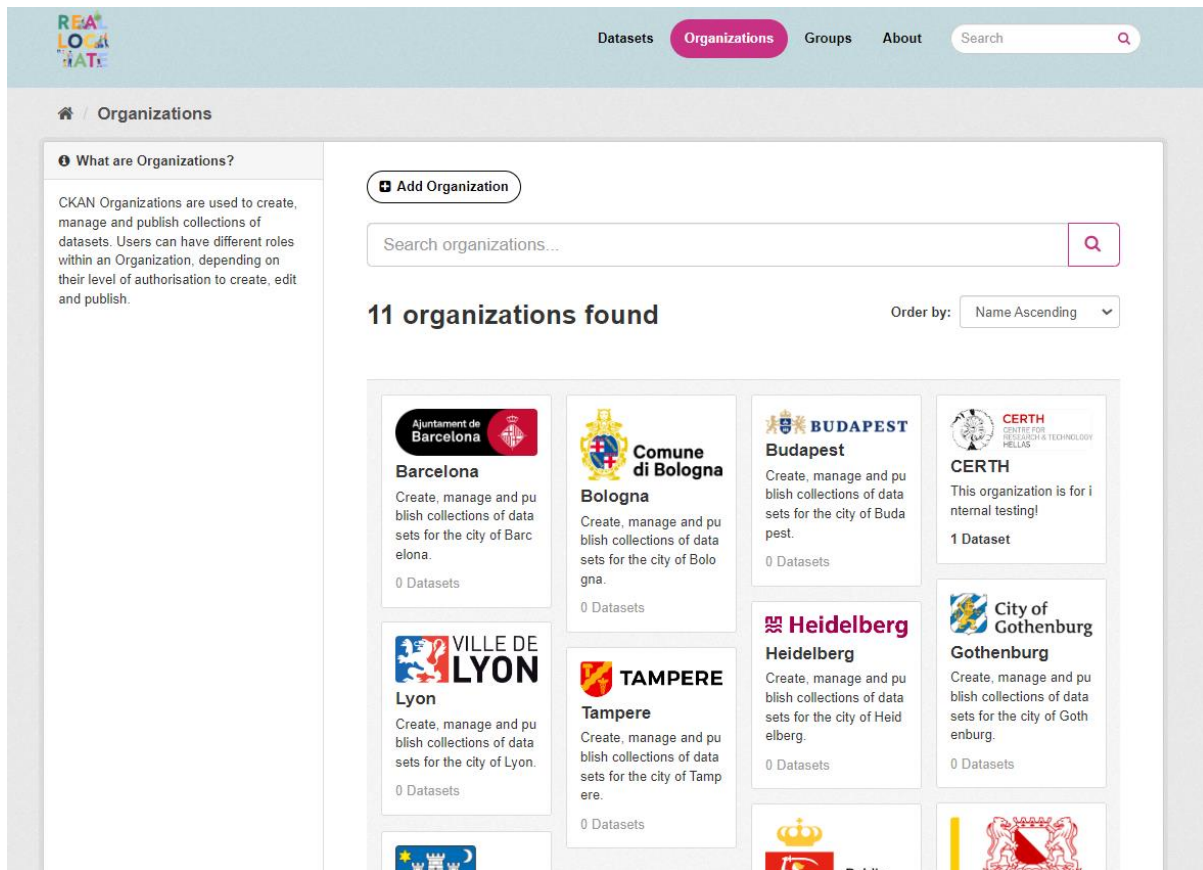
Figure 6 Organization page of the REALLOCATE data platform

Specifically, there are three distinct roles, Member, Editor and Admin, each one having a specific set of responsibilities and permissions over the datasets and the organization itself. A description of the roles and their corresponding privileges is outlined as follows:

- **Member**: Individuals of this role have the ability to view (read-only) the private dataset of the organization. This role is very useful for organizations that wish to share sensitive data with selected users, while still maintaining full control over the data editing and management rights.

- **Editor**: A role specifically targeting the content managers of each organization. In addition to the privileges afforded to members, editors are empowered to upload new datasets, modify existing datasets, and also manage their visibility status (assigning datasets as public or private).

- **Admin**: The role that holds the highest level of authority within an organization. Admins can perform all the functions of both members and editors, and in addition, they have the ability to add or remove users, adjust user roles (including other admin users), manage the details of the organizations (title, description, image etc.) and, if

43

necessary, delete the organization entirely. When an organization is created for the first time, the initiating user automatically becomes the first administrator.

This structured assignment of these roles ensures that data security and governance are maintained at all levels throughout the platform. By limiting access to view-only capabilities on unauthorized users, organizations can confidently share datasets and avoid any unwanted and unauthorized edit or misuse of the shared information. This access control system provides a flexible but yet secure framework, which promotes data security but also encourages collaboration within predefined boundaries.

7.6 Groups

Groups are an essential tool for organizing and clustering related datasets, significantly enhancing the discoverability and improving the accessibility of the datasets within the platform. Each dataset can be assigned to one or more groups, based on common themes, topics, or attributes, making the navigation through the repository easier and more efficient, allowing users to quickly find datasets relevant to their needs. Figure 7 provides an illustration of the groups page in the REALLOCATE CKAN data repository, showcasing how datasets are organized into specific groups.

It is crucial to point out the distinction between groups and organizations within CKAN, as these two components serve a different purpose. Groups are solely used for the categorization of datasets based on similarities, without any implication for membership or access control. They do not involve any form of authorization or ownership rights, but they are purely used to improve the user experience and simplify the discovery of datasets falling under common categories. In contrast, organizations represent entities that have ownership and control over the published datasets. Organizations manage access rights, and determine which viewers have the ability to view particular datasets.

Additionally, while groups do not impose any restriction over the dataset access, the datasets within a group are still subject to the access controls assigned by the owning organization. For example, if an organization designates a dataset as private, that dataset will remain hidden from unauthorized individuals, even if it is part of a public group. Moreover, it is worth noting that while datasets can only belong to one organization, the same principle does not hold on groups.

REALLOCATE Deliverable 1.4 Data Management Plan
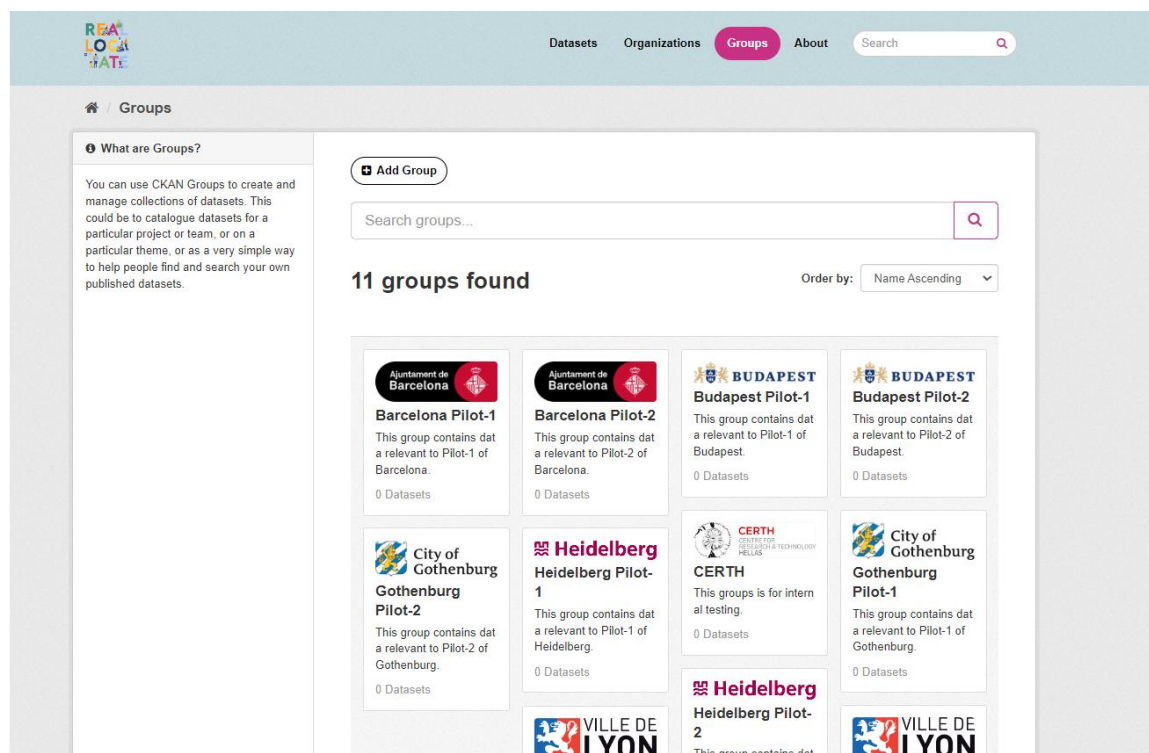
Figure 7 Groups page of the REALLOCATE data platform

CKAN's design allows for greater flexibility by enabling a single dataset to be associated with multiple groups, reflecting its relevance to various themes or categories. This flexibility allows users to explore datasets from different perspectives, and create more accurate groups and ultimately creating a more organized and intuitive browsing experience.

7.7 Application Programming Interface (API)

CKAN is designed with interoperability as a core principle, making it highly adaptable to diverse technological environments and external applications. A feature supporting this flexibility is CKAN's Action API, a versatile and robust tool that extends the full range of platform's capabilities beyond its standard web interface. This powerful API enables developers to interact directly with the platform's data and features, providing this way seamless integration options for the development of external applications.

The Action API provides access to a wide range of CKAN information in a structured, JSON format, making it easy to retrieve lists of datasets, groups, as well as representations of specific datasets and resources. Exposing this information programmatically, an easy and flexible interaction with the CKAN's data repository can be performed, allowing seamless

exchange of information and enabling developers to integrate existing datasets into external applications, dashboards (particularly useful for the needs of REALLOCATE) or analytical tools.

Beyond data retrieval, the Action API also offers full support for managing datasets directly. Specifically, users can perform essential data management tasks such as creating new datasets, updating existing ones, and even deleting datasets that are no longer needed. This functionality is particularly useful and ensures that all datasets remain up-to-date without the need to interact with the CKAN's user interface. Therefore, it is considered highly beneficial for large-scale projects like REALLOCATE, where automatic data management tasks can save significant time by programmatically streamlining the maintenance workflow.

Another key feature of the Action API is its support of advanced search capabilities and access to activity streams. The search functionality is very useful for users or applications that require specific datasets from large repositories. Instead of manually navigating through the web interface, developers can programmatically search for datasets based on detailed search criteria. On the other hand, the activity stream offers detailed insights into changes made over datasets and other platform events such as role changes, group creation, or dataset updates. This feature enables tracking and monitoring activities across the platform, offering a transparent view of actions performed by users, and a version control tool for administrators to audit and ensure that all actions within the platform are properly executed.

7.8 Summary of SHOW's CKAN Methodology

The SHared automation Operating models for Worldwide adoption (SHOW [12]) is a European Union Horizon 2020 research and innovation initiative piloting automated vehicles in urban environments.

REALLOCATE's approach lends elements from SHOW's approach but has certain critical differences. The SHOW project employs a comprehensive DMP to handle data collection, processing, and storage for autonomous public transport fleets across Europe. CKAN is utilized to manage the upload and storage of historical data, allowing participating sites to upload datasets in CSV format. This data is then fed into a MongoDB database for further analysis and KPI calculation. The platform ensures standardization by using predefined data models, which helps maintain consistency across different fleets and manufacturers. KPIs, such as vehicle speed, passenger count, and hard braking events, are calculated and visualized on the publicly accessible SHOW Dashboard. Data exchanges and workflows are secured through APIs and Keycloak authentication.

In addition to CSV file uploads, SHOW facilitates real-time data collection through MQTT and Kafka brokers by directly uploading to the database with the help of a dedicated API, supporting real-time KPI calculation. The collected data is crucial not only for monitoring fleet performance but also for conducting impact assessments and deploying new services such as Estimated Time of Arrival (ETA) [13].

Comparison to REALLOCATE's CKAN Methodology

In comparison, **REALLOCATE's CKAN** system goes beyond CSV files by handling multimodal data formats, such as geospatial data (GIS), sensor data, surveys and time-series data. This makes it more versatile in managing different types of urban mobility data, including data from traffic sensors, public transport, questionnaires and environmental sources. Additionally, great care was taken in making the platform more intuitive for usage by the participating non-technical city representative users, ensuring that they can interact with the system easily.

To enhance user-friendliness, REALLOCATE's CKAN integrates data visualisation extensions, allowing users to preview and analyse data directly within the platform while instantaneously generating complex graphs, enhancing its utility for stakeholders. Specifically, ckanext-visualize is a powerful tool for easy visualisation that is integrated in the project. This extension allows users to create and view charts directly on the platform, making data analysis more direct and intuitive.

Some of the key features of ckanext-visualize include its drag-and-drop interface for creating a chart by clicking on a column from the list of all available columns and dragging it onto the X or Y axis. The extension also offers intelligent chart type selection depending on the data types of the columns selected, automatically visualising the most appropriate chart. There is also support for customizable color palettes and chart type icons. Accessibility is furthermore enhanced by being mobile responsive ensuring usability across different devices.

REALLOCATE will also use a custom plugin that automatically converts datasets into a standardized format for export to the project's dashboard, ensuring smooth data flow and compatibility across the system. The custom plugin being developed for REALLOCATE's CKAN platform is designed to simplify the ETL process, making it more accessible and user-friendly for the participating cities and especially the non-IT technicians uploading data into the system. By automating the transformation step in the ETL process, the plugin makes sure that data is of the same standardized format eliminating the usually error-prone manual data cleaning from the city authorities. Subsequently, the data is formatted appropriately to apply models tailored to the specific data of each pilot site within the REALLOCATE project.

This enables the calculation of KPIs and the application of data analytics or other related services.
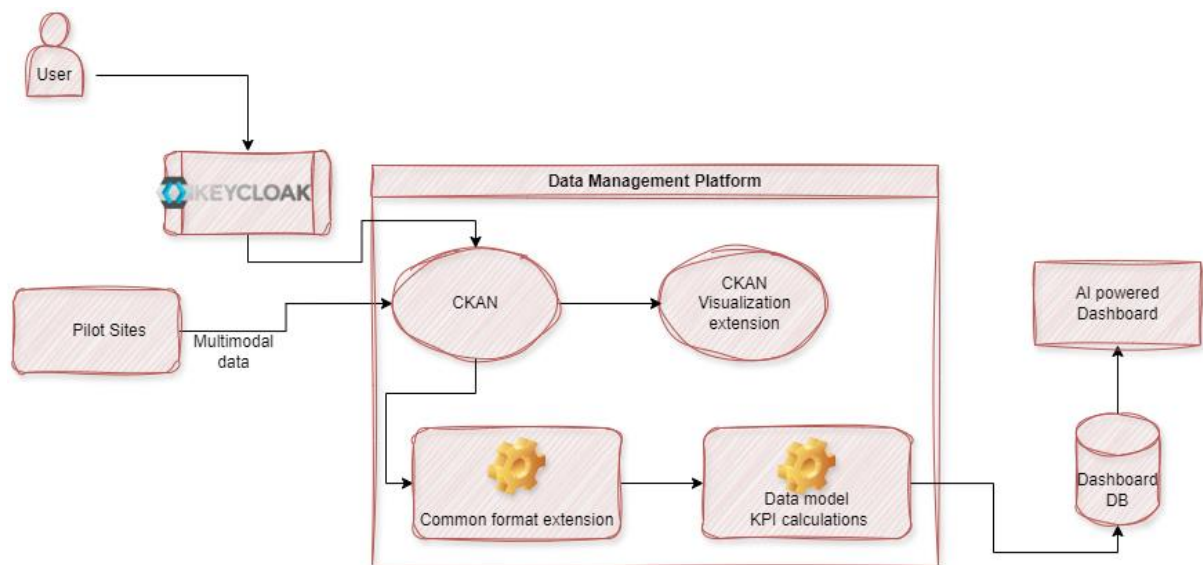
Figure 8 Workflow of REALLOCATE Data Management Platform

Beyond streamlining the data uploads, the plugin can directly pull data from the open data APIs of cities enabling real-time integration of datasets such as traffic flow, public transport usage, air quality and accident data. As such the need for manual data extraction is reduced freeing up time for the cities while ensuring consistency with the project's data framework. The plugin will also be optimized for handling both real-time and historical data, transforming files such as CSV, JSON, XLSX and XML into formats suitable for export to the dashboard. An overview of the role of DMP within the REALLOCATE project is illustrated in Figure 8.

8 Data Security, Storage and Curation

Confidentiality, Integrity and Availability are the 3 core elements adhered to the data security framework that the project partners will structure for the project's data. The aim is to decrease the vulnerability of the information and increase the robustness of the security strategy followed. For the protection of the data that will be collected, stored, created and received or transmitted, secure servers will be used, while controlled access will be provided to specific audience with the use of usernames, encryption and tokenization. Access to individuals will be granted to the project's environment with the concession of the administrators and the PC. Once this process takes place, access is enabled to the user.

CKAN Data Platform

The project uses the **CKAN platform** for managing non-sensitive data. CKAN is an open-source tool that allows the storage, organization, and sharing of data. It makes data easily accessible and searchable for project partners and stakeholders. CKAN also helps with tasks like tagging data with metadata and keeping track of different versions of datasets.

To keep data secure within CKAN, role-based permissions are used to control who can view or edit the data. Only authorized users have access. CKAN also employs **SSL** encryption to protect data while it is being transmitted, and it has logging features to track who accesses and changes the data.

Keycloak User Authentication

The project uses **Keycloak** [14] for user authentication. Keycloak is an open-source system that handles user login and access control. Its features help meet GDPR's principles of data protection by default and by design by making sure that only authorized individuals can access project data, that access is restricted to necessary levels and that personal data is protected.. Keycloak provides additional security features which are expanded upon below.

- **Single Sign-On (SSO)**: Keycloak allows users to sign in once and gain access to multiple systems and applications within the project environment without needing to log in again. This makes it convenient for users to switch between platforms like CKAN and other integrated tools without needing separate credentials.
- **Multi-Factor Authentication (MFA)**: Keycloak supports MFA, which adds an extra layer of security beyond just a password. Users may be required to authenticate via an additional method, such as a one-time code sent to their phone or email, ensuring

that only authorized users can access the platform, even if their password is compromised.

- **Role-Based Access Control (RBAC)**: Keycloak allows administrators to define user roles with specific permissions, ensuring that only users with the correct role can access or modify sensitive data. This helps maintain data security and ensures that each user's access is aligned with their responsibilities within the project.

- **OAuth2 and OpenID Connect Integration**: Keycloak supports industry-standard authentication protocols like **OAuth2** and **OpenID Connect**. These protocols ensure that user authentication is secure and interoperable with other platforms. This feature is also supported by CKAN ensuring secure data access.

- **Social Login Integration**: Keycloak provides the option to integrate social logins, allowing users to sign in using their credentials from social platforms like Google or GitHub. This can simplify the authentication process for external collaborators who may not have internal accounts.

- **Session Management and Timeouts**: Keycloak allows administrators to manage user sessions effectively. This includes setting session timeouts, which automatically log users out after a period of inactivity to prevent unauthorized access. The system also provides the ability to revoke active sessions if a security risk is detected.

- **Token-Based Access**: Keycloak issues secure **access tokens** after a successful login, which can be used by users and systems to interact with protected resources without re-authenticating. These tokens have a limited lifespan, minimizing the risk of unauthorized access in case of token theft or compromise.

- **Audit Logging and Monitoring**: Keycloak keeps logs of all authentication and access events, including successful and failed login attempts, role changes, and session activity. This audit trail is critical for monitoring and responding to potential security incidents, helping project administrators maintain a high level of oversight.

- **Password Policies**: Keycloak enforces customizable password policies, such as complexity requirements and regular password changes, ensuring strong protection against password-related vulnerabilities.

9 Personal data protection

In REALLOCATE, robust measures are implemented to ensure the protection of data and privacy, in accordance with relevant EU laws, particularly Regulation (EU) 2016/679, commonly referred to as the General Data Protection Regulation (GDPR).

Following the principles outlined in the GDPR, the following important factors concerning personal data are taken into account:

1. **Fair and Transparent Processing:** Personal data will be processed fairly, transparently, and in accordance with applicable laws.
2. **Limited Data Collection:** Data collection and usage will be limited to specified, legitimate purposes, avoiding incompatible processing. Personal data may be further processed for archiving, research, or statistical purposes.
3. **Data Minimization:** Personal data will be adequate, relevant, and minimized to what is necessary for the intended purposes.
4. **Data Accuracy:** A commitment to maintaining accurate and up-to-date personal data, is upheld, rectifying any inaccuracies promptly.
5. **Limited Storage Duration:** Personal data will be stored for defined, limited periods, and not retained for longer than necessary for the purposes. Exceptions will be made for archiving, research, and statistical purposes with appropriate safeguards.
6. **Confidentiality and Security:** Personal data will be processed with utmost confidentiality, integrity, and security, utilizing appropriate technical and organizational measures to prevent unauthorized access, loss, or damage.
7. **Responsibility and Accountability:** Stakeholders responsible for collecting, using, and storing data will demonstrate responsibility and accountability, ensuring compliance with these principles.

In alignment with GDPR principles, REALLOCATE places significant emphasis on upholding the rights of data subjects. These rights include:

1. **The right to be informed:** Data subjects are entitled to clear information regarding the collection and utilization of their personal data.
2. **The right of access:** Data subjects have the right to access their personal data.
3. **The right to rectification:** Data subjects have the right to correct any inaccuracies or omissions in their personal data.
4. **The right to erasure:** Data subjects can request the deletion of their personal data when applicable.

5. **The right to restrict processing:** Data subjects can request limitations on how their personal data is processed.

6. **The right to data portability:** Data subjects can obtain and reuse their personal data for various purposes across different services.

7. **The right to object:** Data subjects can object to the processing of their personal data under specific circumstances.

8. **Automated decision making and profiling:** Data subjects have the right to avoid decisions made solely through automated processing, including profiling, when it has legal implications.

In addition to the principles outlined above, it is essential to foster a collective awareness within the Consortium regarding the sensitivity and security of data, as well as its protection. Achieving this shared vision relies on the dedication of Consortium members to the signed CA and the guidelines provided in the DMP.

In compliance with the General Data Protection Regulation (GDPR), all personal data processed within the REALLOCATE project follows strict guidelines for pseudonymization and security. Further details about the consent process for data participants are outlined in the Informed Consent Template (Appendix I).

Any modifications to internal regulations or procedures related to personal data by Consortium partners must be promptly communicated to the coordinator. The coordinator will then relay this information to the relevant granting authority. Furthermore, continuous monitoring of project developments and activities will be conducted to ensure adherence to the established rules and guidelines.

10 Ethical Considerations

The ethical considerations in this section revolve primarily around the safeguarding of personal data, human rights, and the ethical utilization of AI. All project activities should have compliance with the relevant EU and national legal frameworks, ethical principles upholding the highest standards of research integrity [15], fundamental human rights, data protection principles, and, when applicable, the principles outlined in the AI Act.

The project maintains the protection of ethical principles as highest priority with a range of deliverables dedicated to ensuring this protection. Specifically, the following deliverables deal with personal data processing, human rights protection, and ethical principles in relation to AI development and use:

- **D1.4 & D1.5** – The ethical principles governing REALLOCATE are taken into account within the scope of relevant ethical requirements. These considerations are extensively addressed in the forthcoming law and ethics manual, which will be produced as part of D1.4 & D1.5. They delve deeply into the legal, ethical, privacy, and data protection principles and their application to REALLOCATE's activities.

- **D5.1**: Within D5.1, KPIs were introduced to serve as metrics for assessing the impact of interventions. These KPIs are developed collaboratively in partnership with stakeholders, ensuring a comprehensive and tailored evaluation. Importantly, the data collection process will also encompass subjective data, with a specific focus on population groups vulnerable to exclusion. This approach underscores the commitment to addressing ethical concerns and promoting inclusivity.

- **D5.2 & D5.3**: While the primary goal of Task 5.2, as related to D5.2, is to create a rich and adaptive environment using AI techniques for enhanced analytics and multimodal data navigation, it also holds a significant role in ensuring the protection of personal data and ethical considerations. Specifically, within D5.2, there will be efforts to design tools for data collection, including subjective and environmental data, while implementing mechanisms for authentication and pseudonymization of data. These measures are crucial for safeguarding personal data and privacy as the project deals with sensitive information such as surveys. Additionally, the task involves compliance with ethical principles and relevant legislation, ensuring that data processing respects the highest standards of research integrity, human rights, and data protection principles.

- **D5.4**: This deliverable incorporates compliance with EU data protection regulations (e.g., GDPR) and ethical AI guidelines to protect personal data, uphold human rights, and ensure ethically sound AI use. By anonymizing data and using interpretative, explainable AI models, the project mitigates privacy risks and biases, providing urban policymakers with transparent insights. In alignment with the EU's AI Act and ethical AI principles, D5.4 emphasizes transparency, fairness, and inclusiveness, ensuring the AI tools foster equitable decision-making without adversely impacting vulnerable groups. These measures, embedded in the project's framework, demonstrate a commitment to responsible, human-centered AI that supports secure, privacy-conscious urban innovation.

11 Future Work and Conclusions

As the REALLOCATE project progresses, the DMP will continue to evolve to better meet the needs of the project and to ensure compliance with best practices in data management. The initial version (D1.3) provided an essential foundation, outlining the key principles and strategies for managing data in alignment with the European Commission's (EC) guidelines for data sharing and FAIR principles (Findable, Accessible, Interoperable, and Reusable).

The current version (D1.4) introduces several refinements based on the lessons learned and feedback gathered during the initial project phases. These updates include:

- Enhanced procedures for data security
- Improved methods for pseudonymization and anonymization
- A more detailed approach to data sharing and repository management
- Optimization of CKAN functionality for secure, efficient data sharing and visualisation

As the project moves forward, the DMP will continue to develop, particularly as new datasets emerge and policies involve. The final version (D1.5), scheduled for M36, will feature more granular details, especially concerning long-term data preservation, strategies for open-access dissemination, and addressing any newly identified ethical or legal considerations.

The conclusions drawn from this deliverable underscore:

- The critical need for a dynamic and adaptable DMP.
- Secure data management and responsible sharing practices within the REALLOCATE project provide valuable resources and insights to the cities stakeholders, scientific communities and the public.
- Ongoing DMP refinements ensure that data management practices remain robust, compliant and adhere to the highest standards.
- The DMP supports transparency and data reusability, aligning with FAIR principles and GDPR requirements.

References

[1]   CKAN, [Online]. Available: https://docs.ckan.org/en/2.9/user-guide.html.

[2]   E. U. Commission, *Guidelines on FAIR Data Management in Horizon 2020,* 2016.

[3]   CKAN, "The Open Source Data Management System," n.d.. [Online]. Available: https://ckan.org/. [Accessed 31 October 2024].

[4]   Getdkan, "Getdkan," n.d.. [Online]. Available: https://getdkan.org/. [Accessed 31 October 2024].

[5]   S. open data server, "Home | Socrata Open Source | Socrata Labs.," n.d.. [Online]. Available: http://open-source.socrata.com/. [Accessed 31 October 2024].

[6]   J. D. Platform, "Junar Data Platform," n.d.. [Online]. Available: https://junar.com/. [Accessed 31 October 2024].

[7]   Deliverable, "Distributed Dashboard & Centralised visualisation," n.d.. [Online]. Available:
https://reallocatemobility.eu/fileadmin/user_upload/Resources/D5.2_Distributed_Dashb oard___Centralised_Visualisation.pdf. [Accessed 31 October 2024].

[8]   Data.gov, "Data.gov home," n.d.. [Online]. Available: https://data.gov/. [Accessed 31 October 2024].

[9]   T. H. D. Exchange, "Welcome - Humanitarian Data Exchange," n.d.. [Online]. Available: https://data.humdata.org/. [Accessed 31 October 2024].

[10] Energinet, "Home, Energinet," n.d.. [Online]. Available: https://www.energidataservice.dk/. [Accessed 31 October 2024].

[11] E. guide, "Extending guide - CKAN 2.9.11 documentation," n.d.. [Online]. Available: https://docs.ckan.org/en/2.9/extensions/index.html. [Accessed 31 October 2024].

[12] Show, "Shared automation operating models for worldwide adoption," n.d.. [Online]. Available: https://show-project.eu/. [Accessed 31 October 2024].

[13] S. a. O. m. f. W. adoption, "D14.3 Data Management Plan (DMP)," 2022.

[14] C. K. A. N. Documentation, "User Guide - CKAN 2.9.11 Documentation," n.d.. [Online]. Available: https://docs.ckan.org/en/2.9/user-guide.html. [Accessed 31 October 2024].

[15] E. U. Commission, *The European Code of Conduct for Research Integrity,* 2023.

[16] Keycloak, The Linux Foundation, [Online]. Available: https://www.keycloak.org/. [Accessed 31 10 2024].

Appendix I

Informed Consent Template

This project has received funding from the European's Union Horizon Europe research innovation programme under Grant Agreement No. 101103924.

I, [name], (the undersigned), volunteer to participate as a member of the Advisory Board of the project titled "REALLOCATE".

I confirm that (**please tick box as appropriate**):

	Informed consent	
1.	I have read and understood the information about the REALLOCATE project, as provided in the Information Sheet attached with this consent form.	☐
2.	I have been given the opportunity to ask questions about the REALLOCATE project to consider the information and have gotten satisfactory answers.	☐
3.	I understand that my participation is voluntarily, and I can withdraw at any time without giving reasons and that I will not be penalised for withdrawing nor will I be questioned on why I have withdrawn.	☐
4.	In the case of withdrawing, I understand that I should not disclose and/or share any confidential information about REALLOCATE project that I have learned during my participation and according to the clauses in the NDA signed by me and the REALLOCATE Project Coordinator.	☐
5.	I understand the procedures regarding confidentiality and privacy as they have been explained in the Information Sheet attached with this consent form.	☐
6.	I understand that the data collected from our interactions can be used for publications and dissemination as explained in the Information Sheet attached with this consent form.	☐
7.	I understand that the data collected in our interactions will not be re-used for any other purposes than the original purpose of REALLOCATE project as explained in the Information Sheet attached	☐

REALLOCATE Deliverable 1.4 Data Management Plan

	with this consent form.	
8.	I understand that the confidentiality of data collected about me will be preserved as explained in the Information Sheet attached with this consent form.	☐
9.	I understand that my right to request access to any, and all, personal information that I have voluntarily provided as part of my participation, and that I may ask for that information to be rectified and/or amended if it is inaccurate, or request that all personal information that I have provided be deleted.	☐
10.	I understand that any requests for data access, rectification and/or deletion must be done through the project representative (contact details below), that will then forward the request to the REALLOCATE Project coordinator (contact details below) to act upon.	☐
11.	I was informed by the REALLOCATE representative that in case of unexpected findings, the project consortium is obliged to inform: i) The Regulatory and Ethics Advisory Board (…) ii) The Project Coordinator (…) iii) The European Commission via the REALLOCATE Project Officer I understand that the above-mentioned bodies, will decide on the need, means and timing of communicating the findings to relevant stakeholders.	☐
12.	I, an external participant ([name], [company, position]), along with the REALLOCATE team representative, agree to take part in the REALLOCATE study, and to sign and date this informed consent form.	☐

Appendix II

Data Sharing Agreement

1.   Parties

…………….. (hereinafter "**Provider**")

and;

…………….. (hereinafter "**Recipient**")

Provider and Recipient shall hereinafter be referred to jointly as "Parties" and individually as "Party"

WHEREAS the Parties are participants in the project REALLOCATE funded by the European Union (European Commission) through the Horizon Europe – the Framework Programme for Research and Innovation (2021-2027) programme and have signed a grant agreement with the European Commission (the "**Grant Agreement**") and a consortium agreement between the parties participating in the REALLOCATE project (the "**Consortium Agreement**").

WHEREAS the Provider wishes to provide the Recipient, and the Recipient wishes to obtain from the Provider, certain proprietary information and data on terms and conditions set out in this Agreement for the purpose of the research carried out by the Recipient in the ……… project.

NOW THEREFORE THIS AGREEMENT WITNESSETH that in consideration of the premises and covenants set out in this Agreement, the Parties agree as follows:

2.   Definitions

In this Agreement, the following words shall have the following definitions:

"**Agreement**" shall mean this Data Sharing Agreement and its Schedules. In case of contradiction between the provisions and the schedules, the former shall prevail;

 "**Controller**" shall mean, as defined in GDPR Article 7(4), the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of Personal Data;

"**Data Protection Laws**" shall mean Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation, **GDPR**) and any other applicable data protection laws.

"**Effective Date**" shall mean the date upon which the Agreement becomes effective and corresponds to date of the last signature to the Agreement;

"**Personal Data**" or "**Data**" shall mean any information relating to an identified or identifiable natural person(s) considered as personal data under the applicable data protection legislation, especially under the EU General Data Protection Regulation ((EU) 2016/679) included in the Data. The Personal Data disclosed pursuant to this Agreement is further detailed in Schedule A and will be pseudo-anonymized;

"**Dissemination**" shall mean the public disclosure of the Results by any appropriate means (other than resulting from protecting or exploiting the Results), including by scientific publications in any medium;

"**Research Project**" shall mean the research project described in Schedule A hereto. The means in which Personal Data is processed as part of the Research Project have been defined by Recipient;

"**Results**" shall mean any (tangible or intangible) output of the Research Project such as data, knowledge or information — whatever its form or nature, whether it can be protected or not — that is generated in the Research Project as well as any rights attached to it, including intellectual property rights.

3.   **Disclosure of Personal Data**

   3.1.   Provider has originally determined the means and purposes of the collection and processing of Data, thereby constituting the original Controller of Personal Data.

   3.2.   Subject to the terms and conditions of this Agreement, the Provider shall hereby disclose Personal Data to the Recipient.

   3.3.   The Provider legally retains the Personal Data it has collected directly from data subjects based on consent. The Personal Data before being transmitted to the Recipient will be pseudo-anonymized and for the purposes of this agreement.

   3.4.   Upon receiving the Personal Data, Recipient shall constitute an independent Controller of the Personal Data. As an independent Controller, Recipient

represents and warrants it shall process Personal Data pursuant to the Data Protection Laws, as well as good data processing practices, and undertake all rights and obligations relevant to a Controller.

3.5.   Provider and Recipient shall implement continuously appropriate technical and organisational measures to protect Personal Data as required by the applicable Data Protection Laws. Specific technical and organisational measures, including security measures, may be defined in Schedule A.

3.6.   The Provider shall inform the Recipient of requests of the data subjects to exercise their rights as a data subject under the Data Protection Laws. The Recipient shall perform all actions necessary in order to fulfil the rights of the data subject.

3.7.   Provider shall disclose Personal Data in pseudonymized form to Recipient pursuant to Schedule A, and shall not disclose the pseudonymization key to the Recipient. The Recipient shall not use the Personal Data, either alone or in conjunction with any other information, to establish the individual identities of any of the data subjects from whom Personal Data was obtained.

4.   Access Rights to Data and Destroying the Data

4.1.   Provider shall grant Recipient a non-transferable, non-exclusive access right to use the Data solely for the Research Project, and solely for the period commencing on the Effective Date and ending when the Research Project is deemed to have ended as defined in Schedule A, unless terminated earlier in accordance with this Agreement.

4.2.   The Recipient agrees that the Data:

(a)   shall be used only under the Recipient Scientist's direct supervision and only for the purpose of performing the Research Project and for no other purpose;

(b)   shall not be used for any Commercial Purposes;

(c)   shall not be used in research that grants proprietary rights in the Data to a third party; and

(d)   shall not be transferred or disclosed to any third party for any purpose whatsoever without the prior written consent of the Provider.

4.3   Upon termination of this Agreement, the Recipient shall securely destroy the Data and any copies thereof, unless otherwise required by the Provider and/or binding laws and regulations and/or allowed by Data Protection Laws. Irrespective of the abovementioned, the access rights granted to the Recipient by this Agreement are in any case limited to the Research Project only.

5.   Delivery

The manner in which Provider shall deliver Data to Recipient is defined in Schedule A.

6.   Ownership of Results

Results made by the Recipient shall belong to the Recipient, unless they are owned jointly in accordance with the Consortium Agreement. The Provider shall be granted access rights to use the Results in accordance with the Consortium Agreement.

7.   Representations, Warranties and Liability

7.1.   The Data are being provided by the Provider to the Recipient on an "AS IS" basis and the Data is understood to be experimental in nature. Any use of the Data by the Recipient shall be at the sole risk and liability of the Recipient, whether or not the Provider has consented or acquiesced to such use.

7.2.   THE PROVIDER MAKES NO REPRESENTATION OR WARRANTY, WHETHER EXPRESSED OR IMPLIED, WITH RESPECT TO THE DATA AND INFORMATION, INCLUDING ANY REPRESENTATION OR WARRANTY AS TO THE DURABILITY, STORAGE, DISPOSAL, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR TO THE NON-INFRINGEMENT OF THE DATA AND INFORMATION ON THE PROPRIETARY RIGHTS OF A THIRD PARTY

7.3.   The Recipient assumes all liability for damage or loss that may arise from its use of the Data even if the Provider has been advised of the possibility of such damage or loss. The Provider shall not be liable to the Recipient for any loss, claim or demand made by the Recipient, or made against the Recipient by any other party, arising from the use of the Data by the Recipient, except to the extent permitted by law when caused by the gross negligence or wilful misconduct of the Provider.

7.4. Recipient shall be solely liable for its processing of Personal Data, including but not limited to being liable for any losses, expenses and/or damage resulting from Recipient's breach of the applicable Data Protection Laws. Recipient will indemnify and hold Provider harmless for all processing of Personal Data by Recipient or Recipients processors and sub-processors.

8. **Confidentiality**

8.1. During the term of this Agreement, the Recipient shall maintain the confidentiality of any non-public information associated to the Data ("Confidential Information") and prevent any unauthorised access, reproduction, disclosure and/or use of the Data. Confidentiality obligations shall not apply to information that:

(a) was already public at the time of its receipt or has become public thereafter through no breach of this Agreement by the Recipient;

(b) becomes known to the Recipient from a source other than the Provider with no confidentiality obligations as evidenced by prior written records;

(c) was rightfully known to the Recipient prior to disclosure by the Provider as evidenced by prior written records;

(d) was independently developed by the Recipient without the use of the Data;

(e) is approved for release by prior written authorisation by the Provider;

(f) is or becomes public by a law, decree, decision by the authorities, or other similar regulations.

8.2. The Recipient shall use appropriate safeguards to prevent the use or disclosure of the Confidential Information and Data other than as permitted under this Agreement or applicable laws and regulations.

8.3. Notwithstanding the provisions of this Article 8, the Recipient shall maintain the confidentiality of the Personal Data in accordance with the Data Protection Laws.

9. **Dissemination**

9.1. The Recipient shall be entitled to Disseminate the Results in accordance with the terms of the Grant Agreement and the Consortium Agreement. Recipient

REALLOCATE Deliverable 1.4 Data Management Plan

agrees to acknowledge the source of the Data in any Dissemination concerning the Recipient's use of the Data.

9.2. The Recipient must ensure that any Dissemination including the Results shall not include Personal Data and is made in accordance with the provisions of this Agreement.

10. Termination

This Agreement terminates immediately upon the occurrence of any one of the following events:

(a) Mutual agreement between the Parties;

(b) By any of the Parties by means of a prior notice in writing to the other Party within thirty (30) days, in case of serious breach by the other Party of any of the obligations, terms, conditions, clauses or agreements under the present Agreement, always provided that the infringing Party does not remedy such breach within the above thirty (30) day term.

(c) Termination of the Consortium Agreement;

(d) The Recipient becomes bankrupt or insolvent or a receiver is appointed to take possession of the Recipient's business or property or the Recipient has assigned its interest to creditors;

(e) The term of the allowed use defined in Article 4 expires.

11. Notices

All notices, reports, requests, consents and other communications between the Parties pertaining to matters related to this Agreement will be given in writing and delivered by person, registered mail or email, addressed to the Provider Scientist and Recipient Scientist detailed in Schedule A.

Any notice personally delivered or sent by email will be deemed to have been given or received at the time of delivery or transmission. Registered or certified mail will be deemed to have been received on the fifth (5th) day after it was posted.

12. Governing Law and Dispute Resolution

This Agreement shall be governed by and construed under the laws of Greece without reference to its conflict of law rules.

The Parties shall endeavour to settle their disputes amicably. If, and to the extent that, any dispute, controversy or claim has not been settled amicably, it shall be referred to the Courts of the city of Thessaloniki, Greece.

13. Language

The Parties hereby confirm that they both required that this Agreement and all documents and notices in connection therewith are drawn up in English.

14. Schedules

The following Schedules shall form an integral part of this Agreement:

Schedule A. Description of Project and Data

15. Miscellaneous

15.1. The headings used in this Agreement are for convenience and reference only and do not define or limit the scope or affect the interpretation of the provisions of this Agreement.

15.2. No waiver or failure to enforce the strict performance of this Agreement shall be deemed to prevent the Parties from subsequently enforcing their rights. No waiver of a provision of this Agreement will be construed effective unless presented in writing and signed by an authorised representative of the Party granting the waiver or consent. No waiver of a provision of this Agreement will be construed to be a waiver of any subsequent breach of this Agreement.

15.3. The Recipient shall not assign this Agreement, in whole or in part, without the prior written consent of the Provider. This Agreement may only be amended or modified by a written agreement signed by the duly authorised representatives of both Parties.

15.4. The provisions of this Agreement that, by virtue of their nature, are intended to survive any expiration or earlier termination of this Agreement shall do so.

15.5. If any provision of this Agreement is deemed to be invalid or unenforceable, such provision or provisions will be deemed modified to the extent necessary to render the same valid or enforceable, or if such modification is not possible, the remaining terms and provisions of this Agreement will be construed and enforced as if the invalid or unenforceable provision or provisions did not exist.

15.6. Each Party will execute and deliver such further agreements and other documents and take such further actions as the other Party reasonably requests to evidence, carry out or give full force and effect to the intent of this Agreement.

15.7. In case of contradiction between the terms of the Consortium Agreement and the terms of this Agreement, the former shall prevail.

IN WITNESS THEREOF the Provider and the Recipient have caused this Agreement to be executed in duplicate by their respective duly authorised representatives.

_____   _____

SCHEDULE A

DESCRIPTION ON PROJECT AND DATA

1.   Description of Research Project

The REALLOCATE research project aims to reduce transport emissions and improve mobility infrastructure in European cities to achieve climate neutrality. It focuses on rebalancing street space for sustainable modes of transport like walking and cycling, enhancing safety, inclusivity, and urban design. Personal data is processed to evaluate the effectiveness of interventions, ensuring compliance with data protection laws through pseudonymization, which protects individual identities while allowing for data analysis. The project ends when all planned interventions are completed, and the data may be retained for future analysis according to legal requirements.

The purpose of this section is to describe the location of Personal Data, categories of Personal Data disclosed to Recipient and the data subjects to which they pertain and to define the particular technical and organizational measures, if any.

Transfers of Personal Data	Personal Data will be transferred exclusively between participating consortium members in compliance with the General Data Protection Regulation (GDPR). These transfers will facilitate research, data analysis, and collaboration on sustainable urban mobility solutions. All Personal Data, including pseudonymized datasets from surveys and behavioral observations, will be shared only with authorized personnel within the consortium. All transfers will occur over secure, encrypted channels to protect the data during transmission.
Geographic location of storage of Personal Data when received by Recipient	The CKAN platform, used as the central repository for storing and managing the project's data, will host the data within GDPR-compliant regions. These data centers use strict security measures, including data encryption, access control, and regular security audits, to protect the integrity and confidentiality of the stored Personal Data.

REALLOCATE Deliverable 1.4 Data Management Plan

Data Subjects	The data subjects include the horizontal partners and stakeholders from the pilot sites. These encompass a broad range of individuals directly involved in the project's urban mobility interventions, such as policymakers, urban planners, transport engineers, designers, and representatives from civil society organizations. The data subjects also include local stakeholders from the pilot cities, who are engaged in the co-development and testing of sustainable mobility solutions through the Safe & Sustainable Mobility Labs (SSMLs). Additionally, citizens from the pilot cities, particularly those impacted by the urban interventions (e.g., changes in street space allocation and mobility infrastructure), are indirectly involved as data subjects. These citizens contribute through their behavioral patterns and feedback on the urban mobility changes, which is collected and analyzed to assess the effectiveness and inclusivity of the interventions.
Types of Personal Data	The types of personal data processed in the REALLOCATE project include professional contact information (such as names, job titles, and organizational roles) related to pilot representatives and stakeholders from the pilot cities. Data collected for the project may also involve demographic information (e.g., age, gender) of participants in surveys, questionnaires, and workshops. Additionally, behavioral data related to urban mobility patterns, such as the use of transport modes, preferences, and feedback from citizens in pilot areas, may be gathered.
Data security measures	All personal data will be pseudonymized to prevent direct identification of individuals. The project utilizes secure servers located within the European Union, compliant with GDPR, for storing data. Encryption is applied to both data at rest and in transit to safeguard against unauthorized access. Access to sensitive data is restricted to authorized personnel only, with multi-factor authentication used for secure login. Additionally, regular audits and security assessments will be conducted to ensure adherence to data protection protocols, and any sensitive data, such as device or geospatial data, will be anonymized when possible.

2. Quality assurance and validation

Strict quality assurance measures are applied during the processing and preparation of data to be submitted to the recipient. Data undergoes thorough verification and cleaning to ensure accuracy, consistency, and completeness. Personal data is pseudonymized to protect individual identities, and when applicable, anonymized to enhance privacy. Data integrity checks, both automated and manual, are conducted to identify and resolve any issues, ensuring that only high-quality data is processed. Before submission, internal reviews and

audits are performed by authorized personnel to verify compliance with the project's requirements and data protection regulations

3.   Data Delivery

The Provider shall deliver the data to the Recipient in a secure and efficient manner, ensuring the confidentiality and integrity of the data throughout the transfer process. The delivery will be conducted through the use of a designated, secure data management platform, such as CKAN, where the data will be stored and shared with appropriate access controls in place. Only authorized personnel from the Recipient will be granted access to the data, which will be protected through multi-factor authentication. The data will be transferred in a structured and standardized format, ensuring compatibility and ease of use by the Recipient. Additionally, the Provider will ensure that all data is properly anonymized or pseudonymized, as necessary, before delivery to safeguard personal information and comply with data protection regulations.

Appendix III

A. CKAN User Manual

The user manual provides detailed instructions on the full range of the platform's features and functionalities. It is designed to assist all partners in navigating the platform efficiently and managing datasets with ease. The complete user manual is organized and presented in the following sections, ensuring clear guidance throughout the platform.

B. Introduction

Comprehensive Knowledge Archive Network (CKAN) is an open-source data management platform which empowers organizations with tools and functionalities, streamlining the ability to publish, share, store, search and manage datasets. CKAN is the primary data management platform used to fulfil the needs of REALLOCATE, and is the platform where all available data from each pilot site will be stored.

This document aims to assist all involved partners to seamlessly navigate the REALLOCATE CKAN site by presenting all the available functionalities. Specifically, the structure of this document, along with a brief description of the contents of each section, is as follows:

- **Section C**: Introduces the layout and features of the Home Screen, explaining how users can navigate through the data repository, search datasets and understand the statistics of the repository.
- **Section D**: Explains the login process for existing users and what happens when a user cannot gain access to their account. The process of registering users is also explained.
- **Section E**: Provides a step-by-step tutorial on how to upload a new dataset to the platform, including all necessary metadata fields needed to ensure data quality and integrity.
- **Section F**: Describes what organizations are and how datasets relate to them, how to navigate through different organizations and also how to access information about organizations for better understanding.
- **Section G**: Delves into the Groups entities of CKAN, explaining how datasets are clustered together and how to browse, view and create new groups.

It is important to mention that all login credentials will be provided to all partners and no sign-up process is needed. Additionally, following the login process, the only functionality that

should be utilized by the involved parties is the dataset upload feature (Section E.4). The rest of functionalities are provided solely to enhance understanding of the platform's capabilities, and all necessary elements, such as organizations and groups, should be already set up for all parties to use.

C. Home Screen

Upon visiting the REALLOCATE CKAN site[1], users will be greeted with the home screen as presented in Figure_Apx 1. The home screen serves as the gateway to all the data and subsequent features that CKAN has to offer. Specifically, it is divided into two primary components: the Navigation Pane, and the Main Content, allowing users to seamlessly and quickly access all functionalities.
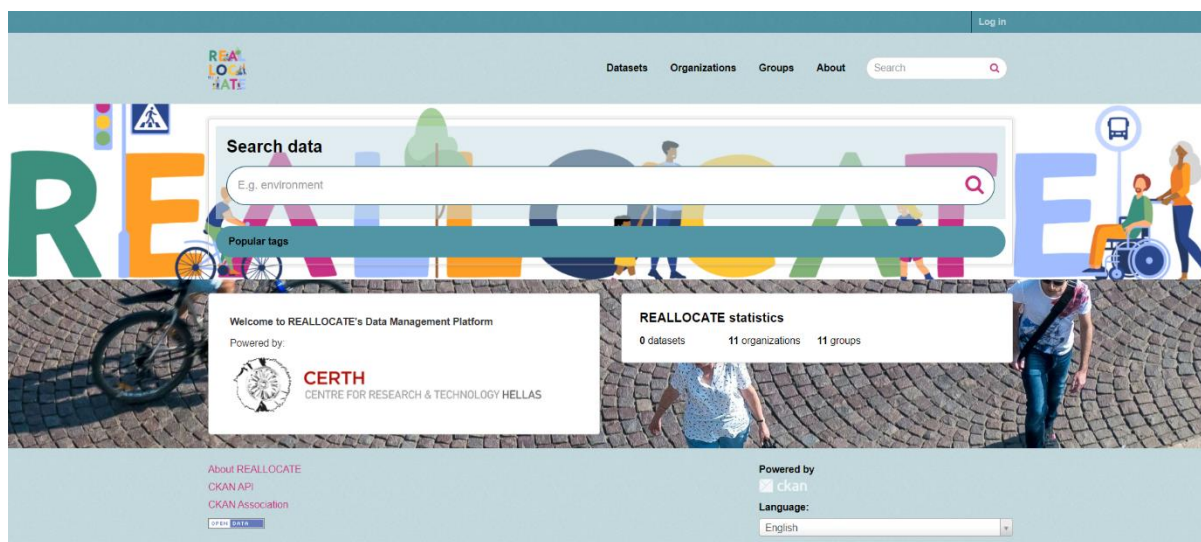
Figure_Apx 1 Home screen of REALLOCATE CKAN site

C.1. Navigation Pane

Located at the top part of the site, the navigation pane constitutes a very important component for the seamless navigation experience throughout the CKAN data repository. The navigation pane is always visible, regardless the part of the site a user is at, boosting this way the ease of use of the platform by providing links to different section of the site that users can access at any time. The sections that the navigation pane consists of are the following:

- **Logo**: prominently displayed at the left side of the navigation page, the Logo serves two purposes. Firstly, it helps users to quickly identify that they have visited the correct site to safely upload data, and secondly, clicking on the logo always redirects users to the home screen, quickly resetting the current navigation.

- **Datasets**: Clicking on Datasets, users are directed to the Datasets page, where they can browse, search and upload datasets. Filtering options are also provided for easier dataset retrieval. Further details provided in Section E.

---

[1] https://reallocate-ckan.iti.gr/

- **Organizations**: The Organizations button directs users to a page listing all existing organizations contributing to the data repository. Users can also browse individual organizations, accessing this way detailed information about the entity, datasets owned, and its members. Further details provided in Section F.

- **Groups**: The Groups button directs users to a page where datasets are organized by different thematic areas, thus datasets with similar topics are clustered together for easier access. Further details provided in Section G.

- **About**: Clicking the About button, users are directed to a page that provides background information about the REALLOCATE project and the purpose of CKAN platform as the data repository.

- **Search bar**: A mechanism used to perform quick and efficient searches for datasets with the use of keywords, particularly useful for users looking to find specific data without the need of further navigation.

- **Log in**: Located at the top right corner of the navigation pane, the log in button directs users to the login page where they can enter their credential and access their account. Further details provided in Section D.

C.2. Main Content

The main content of the REALLOCATE CKAN data repository platform is where users can search and access any available dataset, as well as view statistics that offer a simplified overview of the existing content. To enhance the usability of the site, the main content is divided into two components: the search panel and the statistics panel. A description of how users can utilize both of these components, is as follows:

- **Search panel**: Located at the centre of the site, features a useful search bar where users can input specific keys and find datasets of their interest. A placeholder text ("E.g. environment") is provided, showcasing the correct form of input, and potential user queries. Finally, popular tags (words associated with specific datasets, assigned by users while uploading the respective data) are also displayed below the search bar, offering quick links to commonly searched terms, making in essential for the efficient retrieval of datasets.

- **Statistics panel**: Located at the lower right part of the site, it provides a quick overview of the site's contents. The primary use of this pane is to help users understand the scope and scale of the repository itself, by displaying the total number of datasets, organizations and groups currently available on the platform.

D. Login & Registration Page

The login page is essential part of the REALLOCATE data management platform, enabling users to access their accounts and manage their datasets, while is also assuring the security and integrity of the content existing in the site. The login page can be accessed through the login button of the navigation pane described in Section C.1. This button is always visible, and when a user is logged-in, the respective username is displayed. A visual representation of the login page can be seen in Figure_Apx 2.
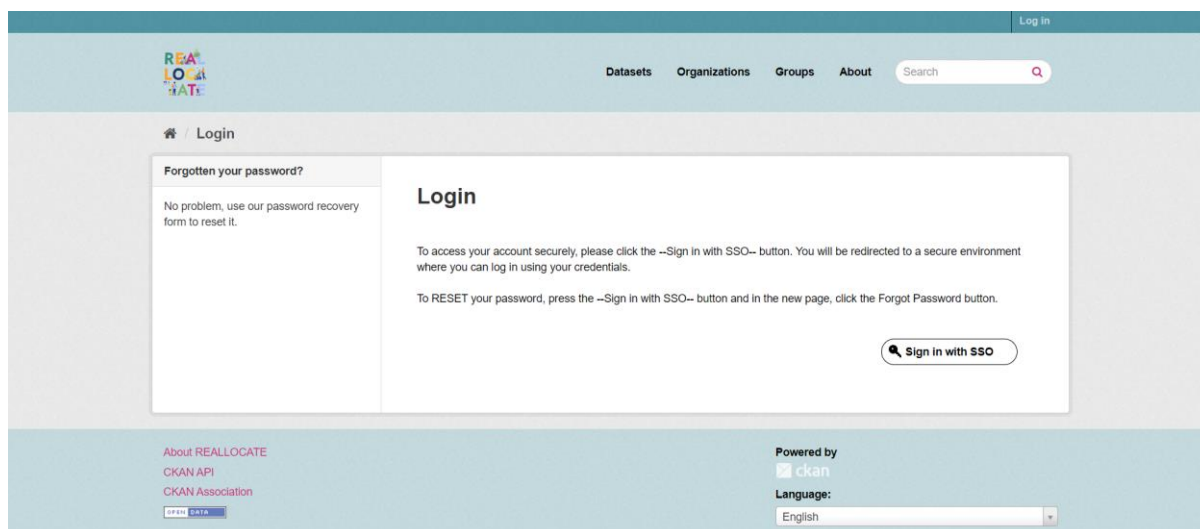
Figure_Apx 2 Login page of REALLOCATE CKAN site
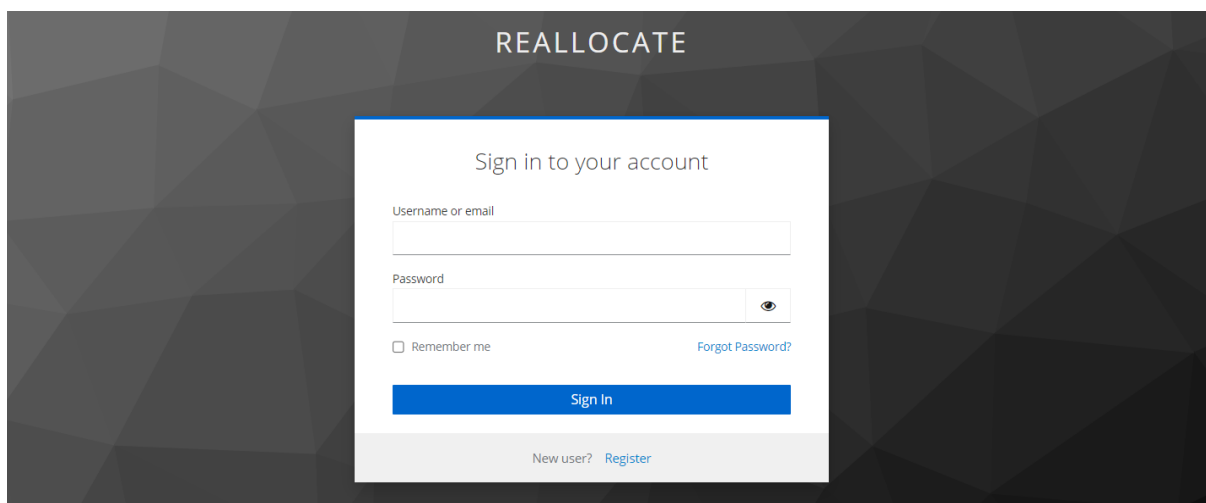
To boost the security of the data management platform, there is only one way to login to the site, which is through the Single-Sign-On (SSO) option. Single Sign On (SSO) is an authentication process that enables users to login to multiple applications (or websites) with one set of credentials. For security purposes, and to protect the integrity and confidentiality of the data managed throughout the data hub platform, all partners are provided with an initial set of credentials created by the site administrator, and can directly log in to the site. Furthermore, SSO offers additional options to users, such as user account creation and account retrieval functionality, while maintaining robust security measures.

It is important to mention that new accounts created through SSO, will initially have no access to any private content existing on the platform. To gain access to such content, users must request permission from the site admin or the owner of the private content they wish to access. This ensures that access is granted securely only to authorized personnel, minimizing the risk of data breaches and maintaining the integrity of the platform.

D.1. Sign-In Form

The login with SSO option is available through the "Sign in with SSO" button on the login page (Figure_Apx 2). Upon clicking this button, users are redirected to the SSO login page (see Figure_Apx 3), which features a form where users can enter their username (or email) and password. Users that already have an active account, can simply enter their credentials and press the "Sign In" button to initiating the authentication process, which, when successful, grants them access to their account and grants entrance to the REALLOCATE CKAN site. Signing in for the first time from accounts created by the site admin requires a password update (see Section D.1.1 for more details). The form also features a "Remember me" checkbox, which, when selected, keeps users logged in on their current device for future visit, without the need to log in again.

In case a user has forgotten their password, or is encountering any issue with the credentials, the "Forgot Password" button can be used (see Section D.2 for more details). Additionally, in case someone wishes to create a new account, the "Register" button located at the bottom of the form should be visited (refer to Section D.3 for more details).
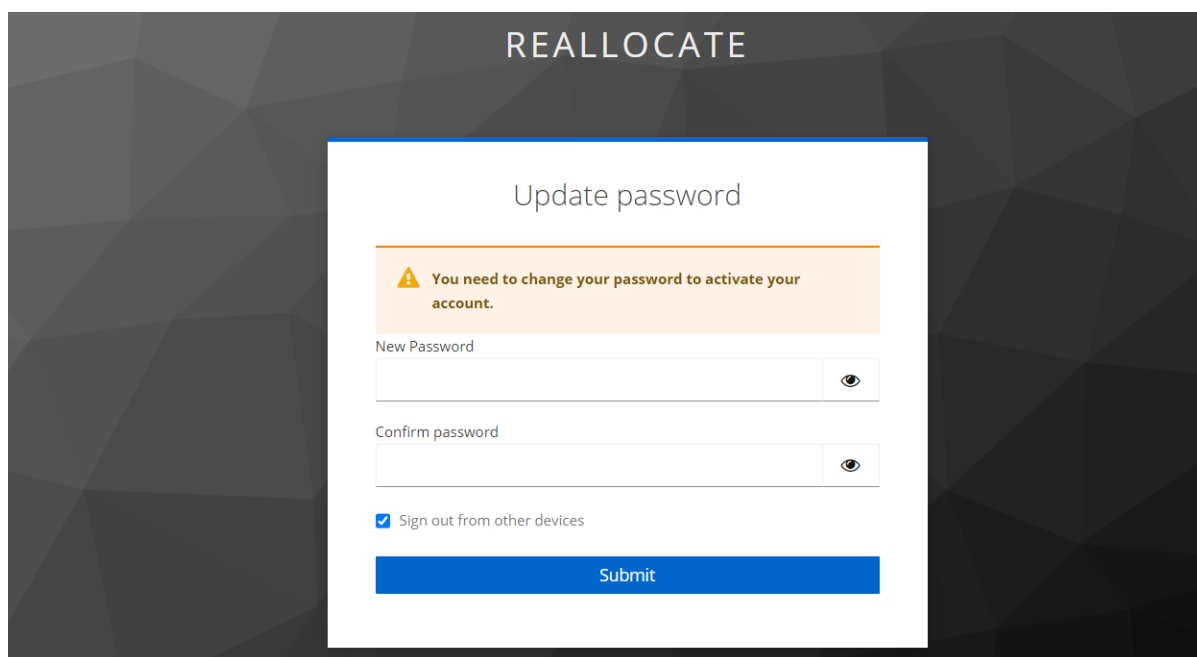
Figure_Apx 3 Sign in with SSO page of REALLOCATE CKAN site

D.1.1. First-Time Sign-In (for Existing Users)

To ensure the security and integrity of the data managed within the data hub, and to streamline the onboarding process, all partners are initially provided with a set of credentials generated by the site administrator. These credentials grant administrative privileges within the respective organizations the users represent.

Upon the first login attempt through the form seen in Figure_Apx 3, users will be prompted to update their password. This initiates the account activation process and users are redirected

to the password update form depicted in Figure_Apx 4. After submitting this form, the password is securely updated, the user's account gets activated and the user automatically gets logged in to the system. It is important to note that, following the password update process, the new password will be known solely by the respective user. This measure ensures that all users have secure access to their accounts from the outset, maintaining the confidentiality and integrity of their credentials. In the event that a user forgets their new password or encounters any difficulties logging into their account, the "Forgot Password" functionality should be initiated (see Section D.2 for more details).

Figure_Apx 4 Password update form of REALLOCATE CKAN site

D.2. Forgot Password Form

Users who have lost/forgotten their password or are experiencing difficulties accessing their account, can easily reset their password by using the "Forgot Your Password" feature. This feature can be accessed via the "Forgot Password" button at the bottom of the SSO login page (Figure_Apx 3). Upon clicking the button, users are redirected to the "Forgot Your Password" form (Figure_Apx 5), where they can input their username or email address to initiate the password reset process.
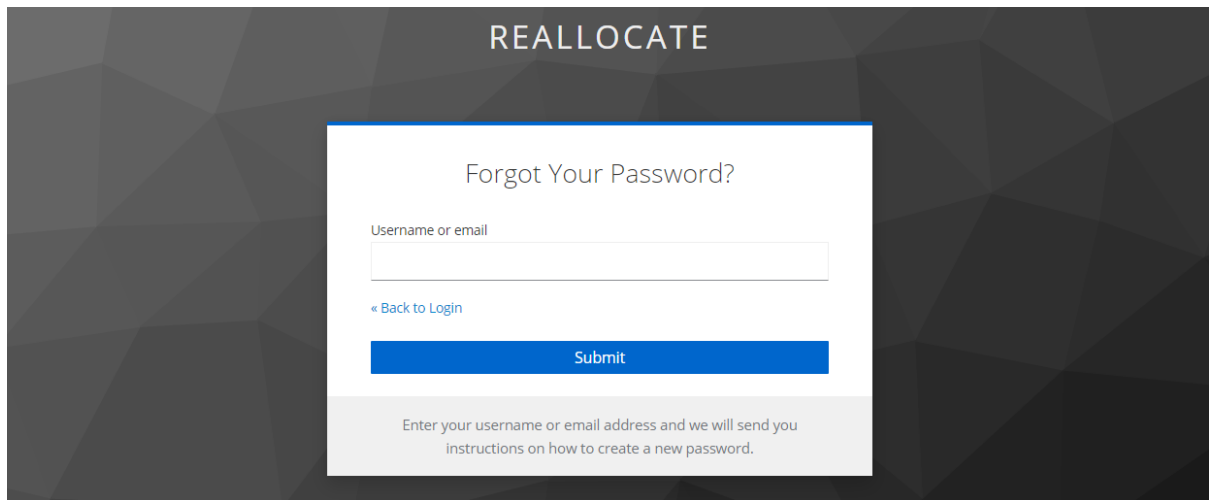
Figure_Apx 5 Forgot your Password page of REALLOCATE CKAN site

After submitting the password reset form, users will receive an email notification as shown in Figure_Apx 6. This email confirms that a reset password reset request has been initiated, and provides instructions on how to create a new password. Specifically, the email will include an expiration link that redirects users to the password update form as depicted in Figure_Apx 4. This link is time-sensitive to enhance security and prevent unauthorized third-party access. Users can follow the link to securely update their password, ensuring this way that all users can safely regain access to their account.
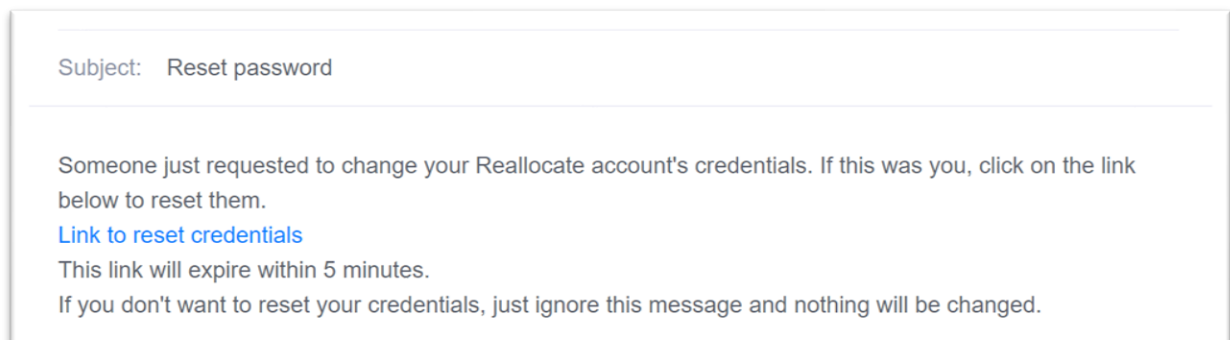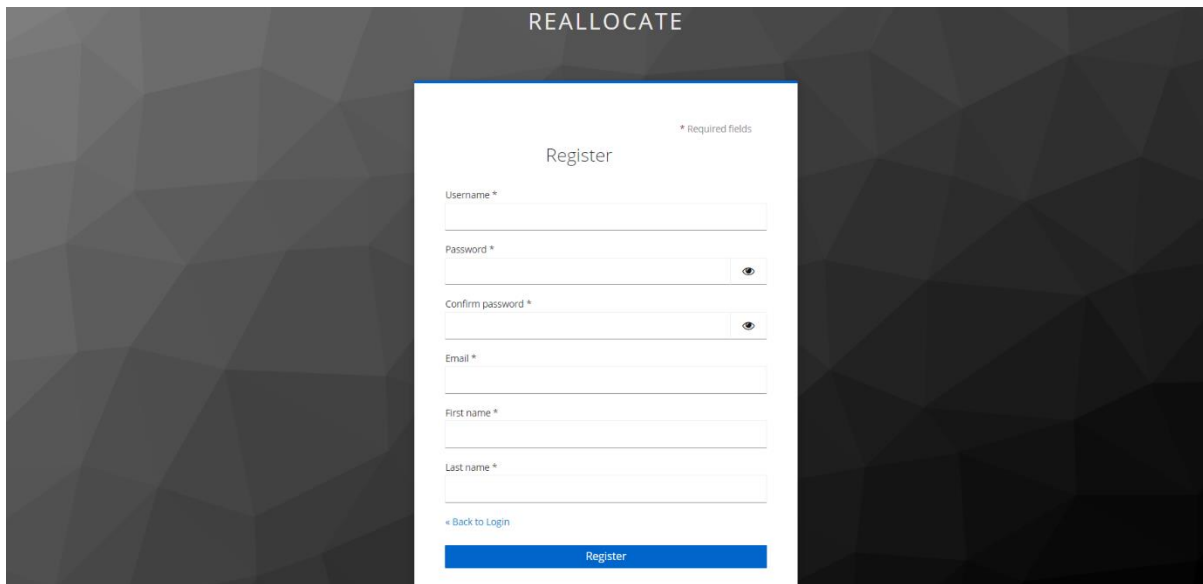
Figure_Apx 6 Reset credentials email

D.3. Registration Form

To create a new account, users must complete the "Register" form, which can be accessed via the "Register" button located at the bottom of the SSO login page (as shown in Figure_Apx 3). Upon pressing the "Register" button, users are redirected to the registration page as depicted in Figure_Apx 7. On this page, the following fields must be filled out in order to successfully create an account:

- **Username**: A unique identifier that will be used for log in purposes.

REALLOCATE Deliverable 1.4 Data Management Plan

- **Password**: A secure password for the account
- **Confirm Password**: field to re-enter the password for accuracy reassurance (must be the same as the one entered in the "Password" field).
- **Email**: A valid email address for account verification/recovery and communication.
- **First/Last name**: The full legal name for identification purposes.

Figure_Apx 7 Registration page of REALLOCATE CKAN site

Once all fields of the registration form have been completed, users must click the "Register" button at the bottom of the form to create the account. When the account is successfully registered, a redirection to the SSO login page is performed, where users can sign in to their new account.

It is important to mention that only authorised individuals will be able to access private content of the repository, preventing this way any unauthorised actions. New accounts will initially have access only to the public content of the site, which is visible to all users, regardless of whether they are logged in. However, these accounts still retain full functionality, including the ability to create organizations or groups and upload datasets. To gain access to any private content on the site, the new user must contact the site admin, or the admin of the organization that owns the private content.

E. Datasets

The Datasets page on the REALLOCATE data hub site serves as the place where users can explore, access and manage available datasets within the repository, or create new ones. Accessed by the "Datasets" button from the navigation pane, this page is a very important component of the CKAN platform, since it provides an organised and user-friendly UI to interact with data. The datasets page can be seen in Figure_Apx 8, while in the following sections, a description of all functionalities provided through the datasets page is provided.
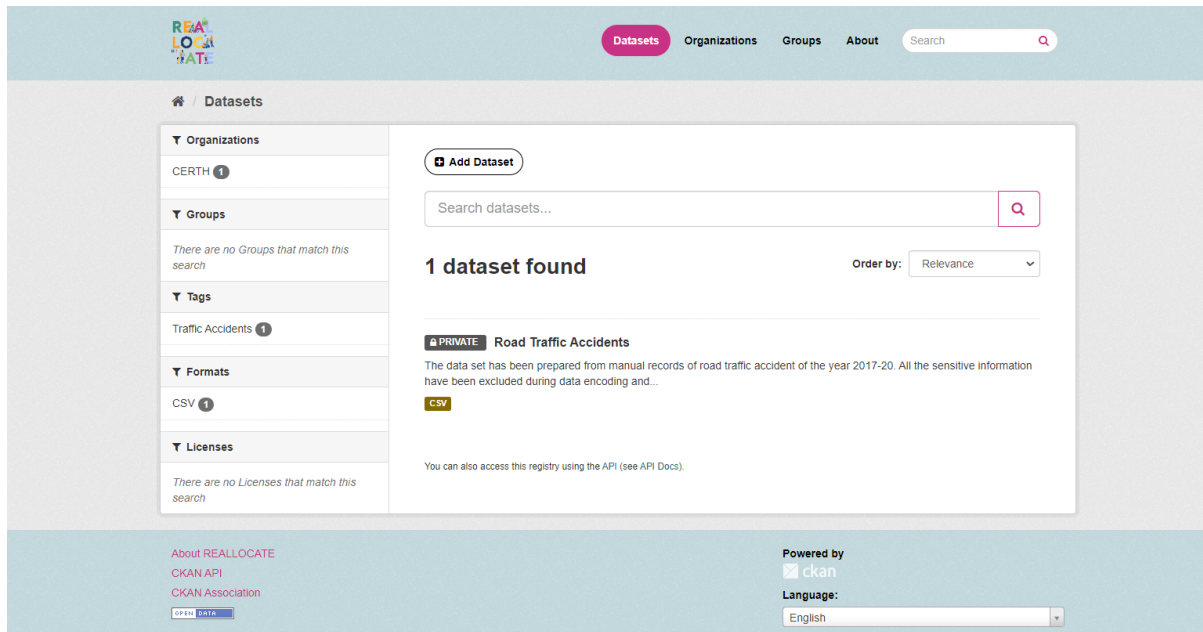
Figure_Apx 8 Datasets page of the REALLOCATE CKAN site

E.1. Search datasets

At the top of the Datasets page, a search bar is located, providing users a quick way to search for datasets by entering their preferred keywords or phrases. This functionality is crucial for locating datasets without the need to browse throughout the whole list of available data. When a search term is queried, a list of datasets appears in descending order based on the relevance to the search term (from most relevant to least relevant). The ordering of datasets can be done by relevance, name ascending, name descending and last modified.

On the provided list of datasets, only the public datasets will be visible to anyone, while private datasets will be visible only to individuals that have the respective access rights. Restrictions to the visibility of datasets are done within organizations, and if a dataset is owned by an organization and is listed as private, only members of that particular organizations will be able to access it.

REALLOCATE Deliverable 1.4 Data Management Plan

Listed datasets are characterised by their unique Title, along with a sort description and the type of the data file. Each dataset is clickable, leading to a detailed view of the dataset. An example of this view is given in Figure_Apx 9.
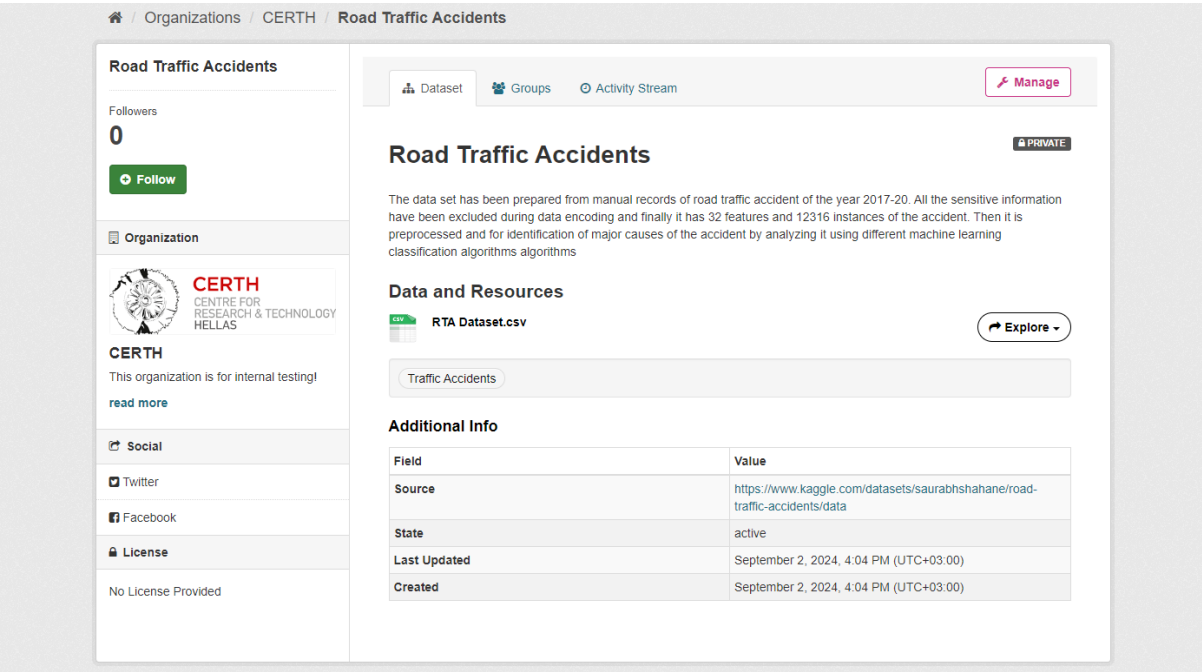
Figure_Apx 9 Viewing datasets in REALLOCATES CKAN site

Specifically, when clicking on a dataset, users are redirected to a page with the following attributes:

- **Dataset**: all dataset details are displayed for the selected data. This page provides a comprehensive view of information such as metadata, data and resources as a linkage to data associated with the selected dataset, and even data previews when applicable (refer to Section E.3 for more details about data previews).
- **Groups**: Provides a listed view of all associated groups that the respective dataset is assigned into. Users can also click on the group to view further details about the group itself. More information about groups is available in Section G.

E.2. Filter datasets

At the left side of the datasets page (Figure_Apx 8), a variety of filtering options is offered, allowing users to narrow down their search preferences even more, based on different criteria. Users can filter datasets by organizations, groups, tags, formats and licenses. When a filter is selected, the pool of filters updates, excluding this way filters that no longer apply in combination with the selected filter. This functionality helps users find datasets that meet specific requirements in an easy and intuitive way.

REALLOCATE Deliverable 1.4 Data Management Plan

E.3. Preview datasets

Within the REALLOCATE CKAN site, users can quickly preview uploaded datasets to gain insights before using them. In order to preview a dataset, the following steps should be performed:

1. Visited the dataset's page through the navigation pane (Figure_Apx 8).
2. Select the dataset you wish to preview (Figure_Apx 9).
3. Press the "Explore" button on the left side of the resource you want to preview.
4. Select the "Preview" option on the drop-down menu

A redirection to the dataset preview page will be performed, and users will be able to preview the selected resource using one of three available preview options:

- **Grid**: Displays the data in a tabular format, allowing easy examination of individual data points (left preview of Figure_Apx 10).
- **Graph**: Visualizes the data through various types of charts and graphs, providing a quick overview of trends and patterns (right preview of Figure_Apx 10).
- **Map**: Geospatially represents data points on a map, only applicable to datasets containing location-based information.
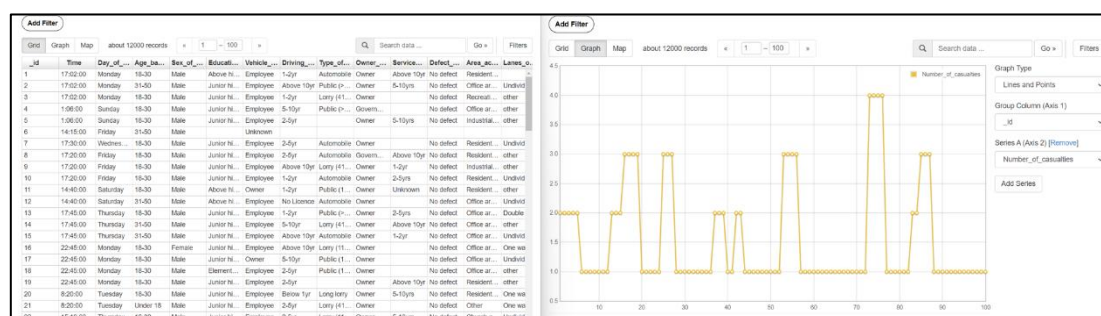
Figure_Apx 10 Preview of datasets in REALLOCATE CKAN site

E.4. Upload datasets

The primary use of the REALLOCATE CKAN site is to provide users a way to safely share and manage their datasets with multiple other authorised individuals. By accessing the Datasets page through the navigation page, users are greeted with the page illustrated in Figure_Apx 8. When a user is logged in with valid credentials, an "Add Dataset" button is visible above the list of available datasets, while on the other hand, users without credentials can only view the available public datasets and do not have the ability to add new ones.

E.4.1.   Create dataset

By clicking on the "Add Dataset" button, users are redirected to the "*Create Dataset"* page as seen in Figure_Apx 11, where all information regarding the data can be initialised. Specifically, users can add the following information:

- **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.

It is important to mention that the "Title" field is required and users must include it, since it stands as a unique identifier among all datasets within the data repository. The rest of the fields are not mandatory, but is highly advisable to provide as many as possible to ensure the appropriate use of the respective dataset from other users, as well as the seamless integration of the dataset with the rest functionalities provided by CKAN (comprehensive previews and searchability through filters). After completing all applicable fields for the

---

* Required

respective dataset, users can move to the next and final form by clicking on the "Next: Add Data" button at the bottom of the page, in order to complete the process of uploading the dataset. This form is analysed in Section E.4.2.

Figure_Apx 11 "Create Dataset" page of the REALLOCATE CKAN site

E.4.2. Add data

The final stage of uploading a dataset to the REALLOCATE CKAN site, is the "Add data" form. Users after initializing all relevant information about the dataset on the "Create dataset" form (Section E.4.1), they can upload the actual data file as seen in Figure_Apx 12. Specifically, this form consists of the following fields:

- **Data**: Either Upload or Link. By choosing upload, file browser will appear for users to upload a data file from their local device, while choosing link, users can specify a link pointing to the source of the data file.
- **Name**: A name for this specific resource.
- **Description**: A short description for this specific resource.
- **Format**: The file format of this specific resource (typically added automatically by CKAN and can be skipped).

It is important to mention that the fields "Name" and "Description" of the "Add data" form are different from the fields "Title" and "Description" of the "Create dataset" form of Section E.4.1. All fields of the "Create dataset" form refer to information regarding the dataset, while fields

of the "Add data" form refer to information regarding a specific data source (file containing data). A dataset can contain more than one files of the same or different file format.

After completing all the relevant fields, users can choose one of the following actions:

- **Previous**: user gets redirected to the "Create dataset" page (Figure_Apx 11) for possible changes on the given fields.
- **Save & add another**: the current data are saved and the "Add data" form resets for the user to add another file.
- **Finish**: the current data are saved and the process is completed.

Figure_Apx 12 "Add Data" page of the REALLOCATE CKAN site

E.4.3.   Assign data to groups

Cities with more than one pilot sites should assign their datasets to the respective group, embracing this way an organized and intuitive platform, where all involved partners can directly target and find the data relevant to a specific site. For the cities with two pilot sites, a group has already been created for each respective pilot (see Figure_Apx 22), while for cities with only one pilot, the clustering of the data can be achieved through the organization, attributed during the "Create dataset" process (Section E.4.1). If additional groups are needed, specific instructions on how to add new groups are included in Section G.1.

To assign a dataset into a group, do the following steps:

- **Step 1**: Select the dataset you uploaded (e.g. Figure_Apx 9)

REALLOCATE Deliverable 1.4 Data Management Plan

- **Step 2**: Select the Groups pane at the upper side of the page (Figure_Apx 13). The groups that this specific dataset is already assigned to will be shown.
- **Step 3**: Press the drop-down menu and select the Group you want your dataset to be added to.
- **Step 4**: Press the "Add to group" button and the dataset will be added to the selected group.
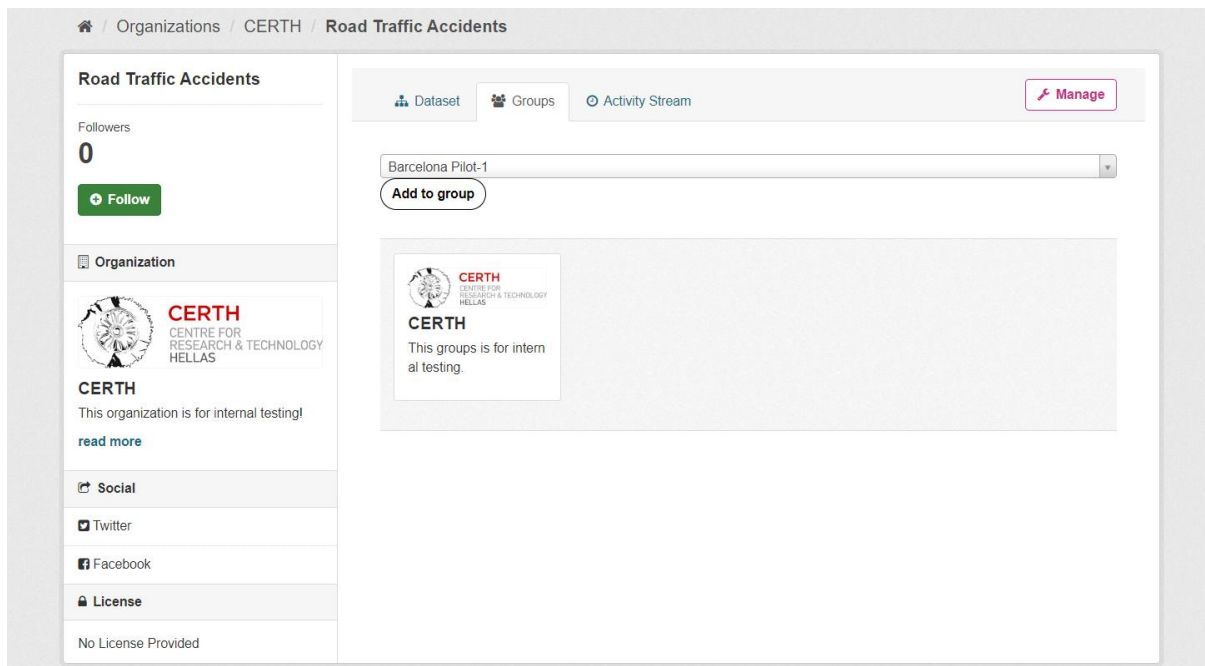
Figure_Apx 13 Groups pane of a specific dataset

In case a dataset should be removed from a group, steps 1 and 2 remain the same, while as a final step users should hover the cursor over the group the dataset was falsely assigned to, and a "remove" button appears. By pressing this button the dataset will successfully be removed from that specific group.

E.4.4.   Simplified instructions

This section is focused on providing easy and direct step by step instructions to upload data to the REALLOCATE CKAN site. Each step may also include references to figures and the respective sections that provided further details. The steps to upload a dataset are as follows:

- **Step 1**: Visit the REALLOCATE CKAN site[2]. You will be greeted with the home screen as seen in Figure_Apx 1.

---

[2] https://reallocate-ckan.iti.gr/

- **Step 2**: Login in with your credentials. Refer to Section D for instruction on how to login.

- **Step 3**: Select the "Datasets" button at the navigation pane of the home screen.

- **Step 4**: Click the "Add Dataset" button at the upper part of the page (Figure_Apx 8).

- **Step 5**: Complete the "Create dataset" form (Figure_Apx 11). Details about all fields in the form are provided In Section E.4.1. Note: do NOT forget to select your organization on this form.

- **Step 6**: Press the "Next: Add Data" button at the lower part of the page.

- **Step 7**: Complete the "Add data" form (Figure_Apx 12). Details about all fields in the form are provided In Section E.4.2.

- **Step 8**: Press the "Finish" button at the lower part of the page to complete the process.

- **Step 9**: Assign the newly created dataset into a group (applicable only to cities with more than one pilots). For further information refer to Section E.4.3.

E.5. Edit dataset

The RALLOCATE CKAN site provides robust editing functionalities that allow users to update and manage their datasets at any point of time. The editing feature is crucial since users have the ability to keep their resources up-to-data and accurate. This section focuses on what fields of the data can be edited, and thus, constitutes a guide on how to edit datasets and their respective fields. It is important to mention that users can only edit datasets created by themselves, or create by organizations that they are a member of.

To edit a dataset, follow the next steps:

- **Step 1**: Access the dataset's page of the data you want to edit as seen in Figure_Apx 9 (can be access either by directly searching the title of the dataset, or by navigating to it through the "Datasets" page from the navigation pane).

- **Step 2**: Press the "Manage" button at the upper part of the page.

- **Step 3**: The "Edit metadata" page appears (Figure_Apx 14), with similar fields as the ones seen in Section E.4.1. Users can also add new resources or edit existing ones on this particular dataset by pressing the "Resources" tab right next to the "Edit metadata" (Figure_Apx 15). To edit an existing data source, select the file you want to edit, and a form similar to Figure_Apx 12 will appear.

- **Step 4**: when all relevant fields are edited, press "Update dataset" at the bottom part of the page to finish the process.

Figure_Apx 14 "Edit metadata" page of the REALLOCATE CKAN site

Figure_Apx 15 "Edit Resources" page of the REALLOCATE CKAN site

REALLOCATE Deliverable 1.4 Data Management Plan

F. Organizations

Organizations are the primary way of managing access to datasets within the REALLOCATE CKAN site. When a dataset is uploaded to the platform, it gets associated to strictly one organization, meaning that access to particular datasets is managed exclusively by its owning organizations. This section provides a detailed overview of how the organizations page works, along with all its functionalities.

The organizations page can be accessed by pressing the "Organizations" button from the navigation pane. A preview of the organizations page can be seen in Figure_Apx 16. A list of all available organization on the platform will appear. For easier access to a desired organization, a search bar is provided, as well as ordering options (organizations listed in ascending or descending order by their name).
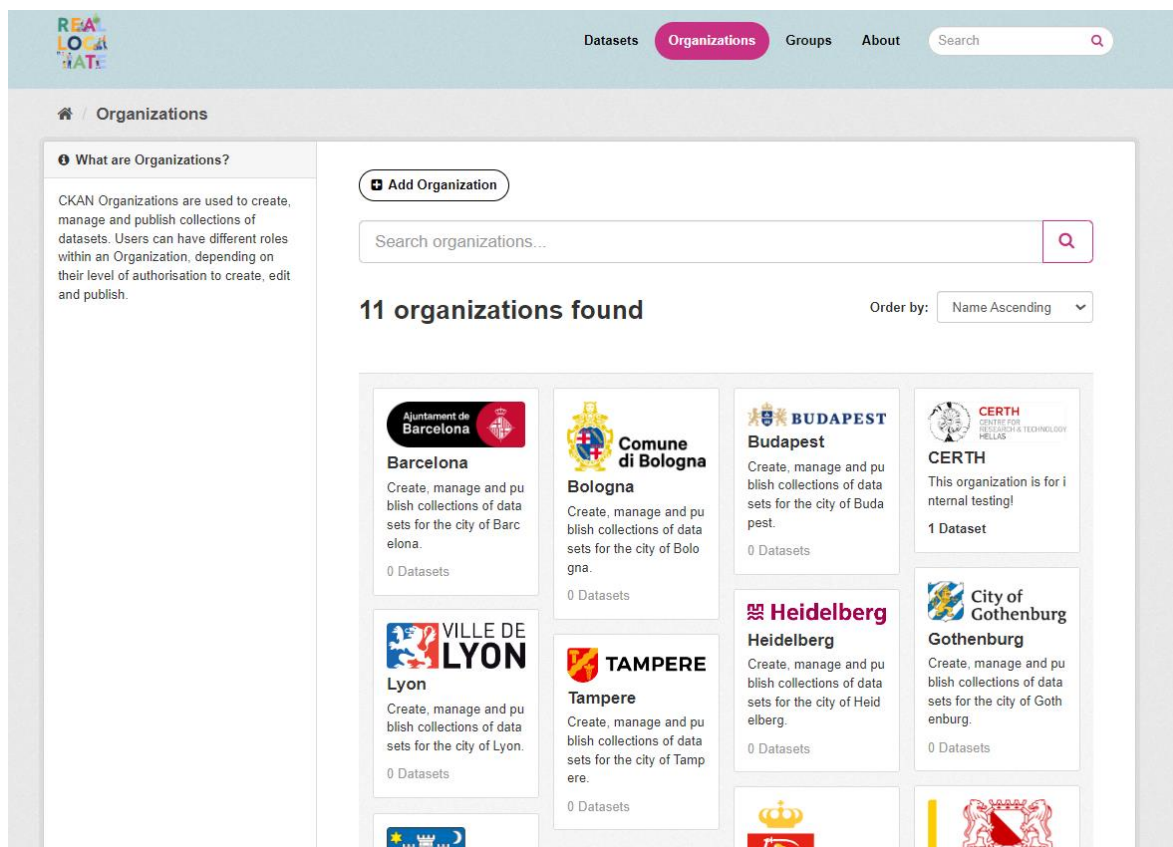
Figure_Apx 16 Organizations page of the REALLOCATE CKAN site

By selecting an organization, users are redirected to the organization detail page as seen in Figure_Apx 17. A list of all datasets owned by the selected organization is displayed. Users can also view the "About" section that provides a description of the organization.

On the left side of the page, statistics regarding the followers and the number of datasets owned by the organizations are displayed. Additionally, filtering options are available for easier access to the desired dataset.
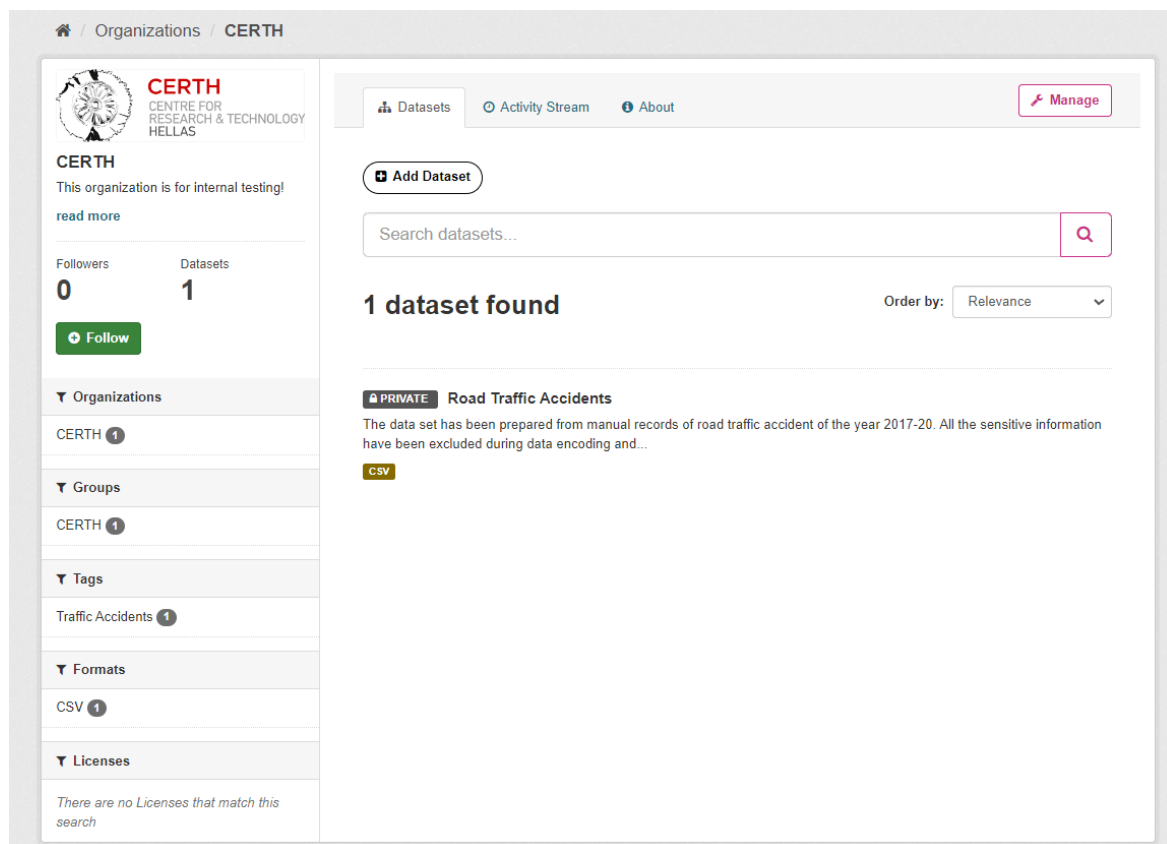
Figure_Apx 17 Organizations detail page of the REALLOCATE CKAN site

F.1. Create organizations

Logged-in users have the ability to create new organizations through the "Add Organization" button at the upper part of the organizations page as seen in Figure_Apx 16. When clicked, a redirection is performed and the page seen in Figure_Apx 18 appears, where users can input a title, description, and an image or URL with a logo (optional) for the organization. Pressing the "Create organization" button at the lower part of the page, creates the new organization. The user that creates an organization automatically becomes the first administrator (admin) and holds the highest level of authority within the organization and can manage the access of other individual users. Section F.2 describes the process of managing access of users within organizations.

Figure_Apx 18 Create organization page of the REALLOCATE CKAN site

F.2. Manage access

Users that create an organization automatically become admins of that respective organization, and thus, have the highest level of authority within that organization. Admin users have access to the admin panel, as seen in Figure_Apx 19. To access the admin panel, visit the organizations page and select your organization (Figure_Apx 17), press the "Manage" button at the upper right side of the page and select the "Members" pane (Figure_Apx 19).

Figure_Apx 19 "admin" page of the organization

Through the admin panel, admin users can manage the access of users within the organizations. Within organizations, individuals get assigned into one of the three distinct

roles: member, editor or admin. 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.

To update the role of an existing member within your organization, navigate to the admin panel and select the "Members" pane. Locate the user whose role you wish to modify, then click on the "gear" icon situated on the right side of their username. This action will open a new page (Figure_Apx 20), where you can adjust the role of the selected user through a drop-down menu.
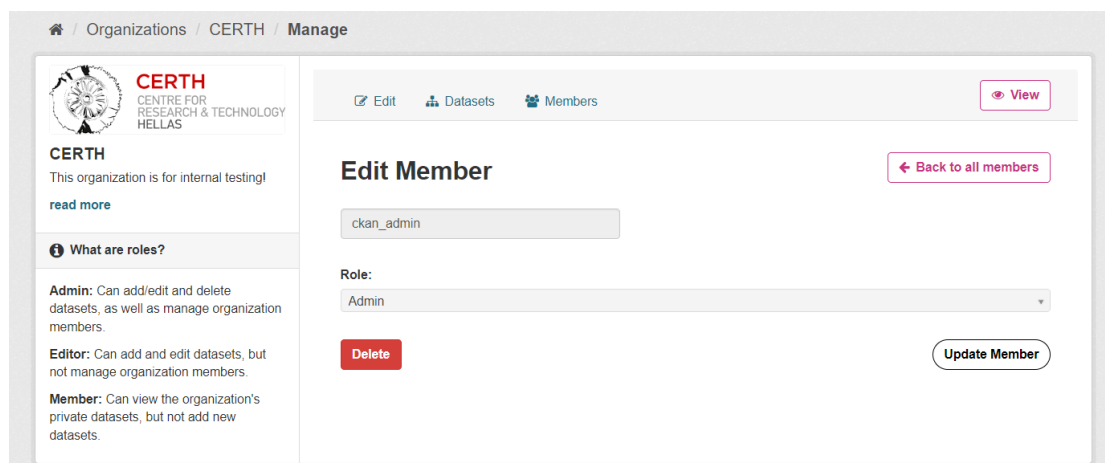
Figure_Apx 20 Edit role of members in organization

F.3. Add/Remove members

Admin users within the organization have the ability to manage access and determine who can be included in the organization. To add or remove a member, admin users should visit the admin panel page by performing the following steps:
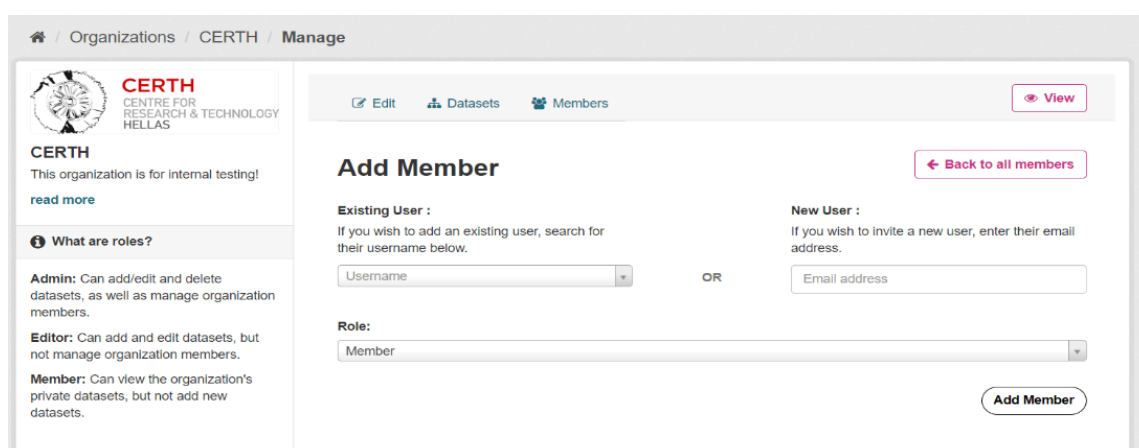
1. Navigate to the "Organizations" page through the navigation pane (Figure_Apx 16)
2. Select your organization. Note: You must have admin privileges (Figure_Apx 17)
3. Click the "Manage" button located at the upper-right corner of the page to access admin panel

REALLOCATE Deliverable 1.4 Data Management Plan

4. Select the "Members" pane on the upper part of the page (Figure_Apx 19)

After completing these steps, a page displaying all current members of the organization will appear. From this page, admins can add new members or removing existing ones. Detailed instructions for both these functionalities are provided in the following sections.

F.3.1.   Add new members

To add a new member to the organization, click the "Add Member" button located at the upper part of the admin panel (Figure_Apx 19). A new page will appear as seen in Figure_Apx 21, which allows admin users to either add a member through a drop-down menu (for members that already have an account in the REALLOCATE CKAN site), or by entering their email address if they wish to add a member that does not already hold an account within the site. After identifying the user, the admin must assign a role (Member, Editor or Admin) though a drop-down menu. This role indicates the authority that the new user will have within the organization. For a detailed explanation of each role's privileges, refer to Section F.2.

Figure_Apx 21 Add new members within the organization

F.3.2.   Remove existing members

Admin users have the authority to remove any member from the particular organization that they manage. To remove an existing member of the organization, navigate to the admin panel (Figure_Apx 19), select the "Members" pane, and locate the user you wish to remove. On the right side if the user's name, click the "X" button to initiate the removal process. Please note that this action is irreversible, and the user will lose all access to the organization's resources immediately upon removal. The user can be added again by following the process described in F.3.1.
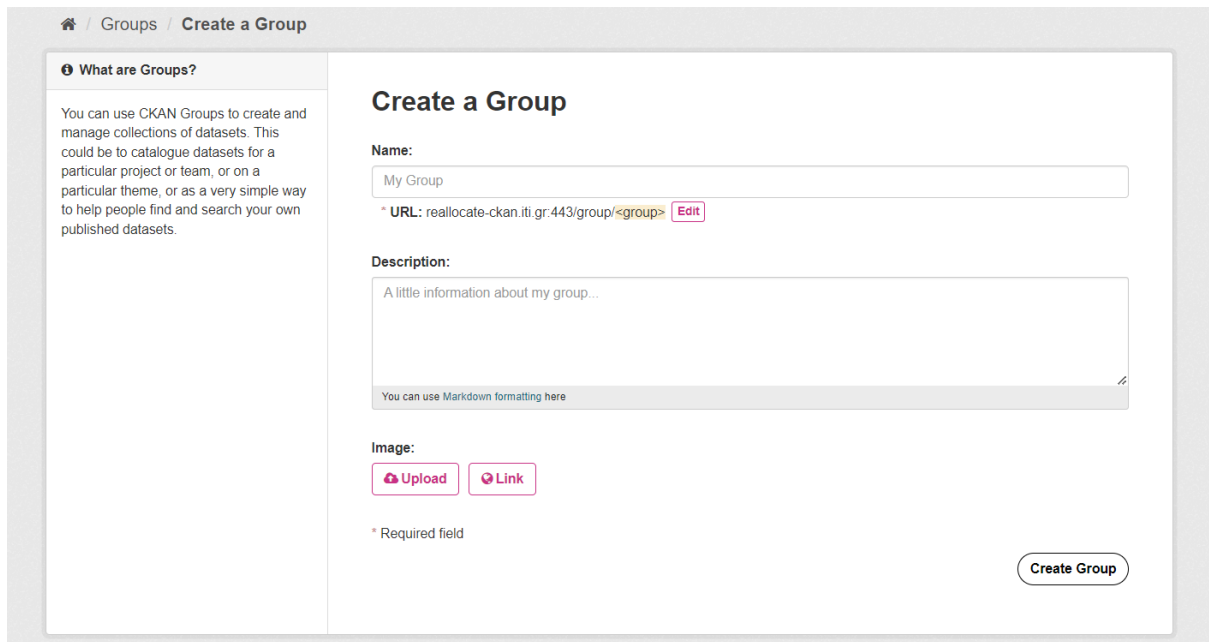
G. Groups

Groups are used to cluster similar datasets together, boosting the ease of discovery and accessibility. The Groups page can be accessed by pressing the "Groups" button of the navigation pane, and users are redirected to the page illustrated in Figure_Apx 22, where a list of all available groups is presented.
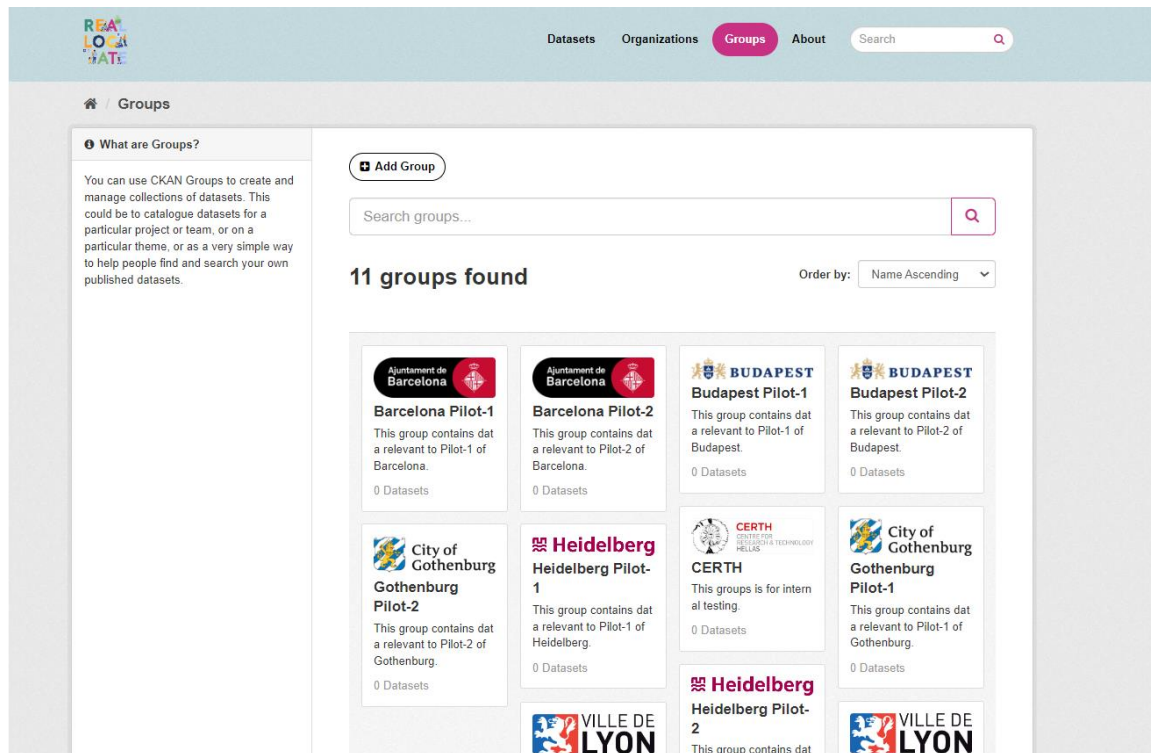
Figure_Apx 22 Groups page of the REALLOCATE CKAN site

G.1. Create groups

Logged in users are able to create their own groups to cluster their datasets (for non-logged in users this functionality is not available). To create a new group, press the "Add Group" button at the upper part of the groups page (Figure_Apx 22). A redirection to the form seen in Figure_Apx 23 will be made, where users can insert the name and description of the group they want to create, as well as an image or a URL for the logo. To finish the process, press the "Create group" button at the bottom of the page. Additional information regarding the group and its associated datasets, can be seen when selecting a group (Figure_Apx 24). This information can also be edited by pressing the "Manage" button at the upper right side of the page (edit group name, description, logo and members).

REALLOCATE Deliverable 1.4 Data Management Plan
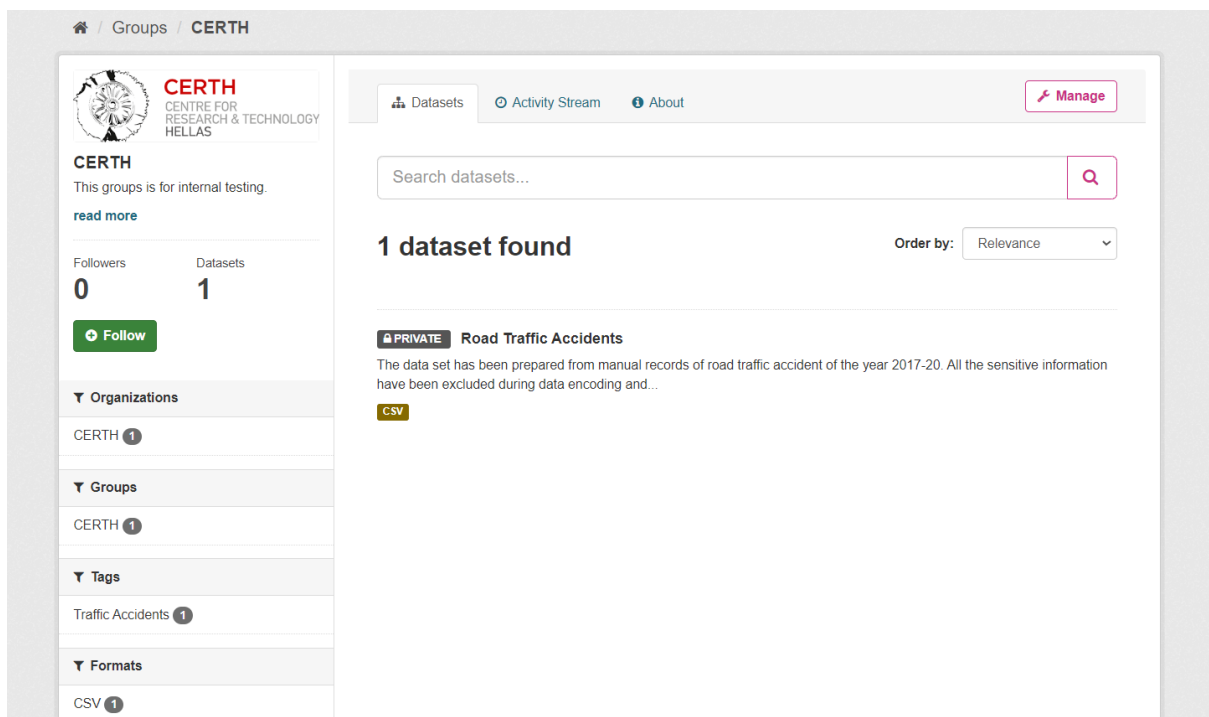
Figure_Apx 23 Create groups in the REALLOCATE CKAN site

Figure_Apx 24 Viewing a group of the REALLOCATE CKAN site

REALLOCATE Deliverable 1.4 Data Management Plan