Data Management Plan

Deliverable number: D1.3

Date: 31/10/2023

Version: v1

Name of work package	WP1 – Project coordination
Dissemination level	Public
Due delivery date:	31/10/2023
Actual delivery date:	31/10/2023
Lead beneficiary:	CERTH
Contributing beneficiaries:	-

Abstract

The current deliverable will outline the comprehensive data management strategy of the project. It will detail the approach towards data collection, curation, storage, preservation, and sharing, while ensuring compliance with the European GDPR (General Data Protection Regulation) and associated national legislations. Adhering to the FAIR data management principles, the deliverable will focus on ensuring data findability, accessibility, interoperability, and reusability. Moreover, the plan will highlight 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. It is important to note that two subsequent versions of this deliverable, D1.4 and D1.5, are scheduled to be released on M18 and M36 respectively, to accommodate potential refinements and updates to the data management strategy as the project progresses.

Table of Contents

List of tables

List of figures

List of acronyms

Acronym	Explanation
AI	Artificial Intelligence
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
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

About REALLOCATE

REALLOCATE transforms streets into inclusive, green, safe and future-proof urban spaces, where communities live and thrive. The project enables researchers, mobility experts, urban planners and local citizens to collectively re-imagine our cities and redesign how we move from one place to another.

1 Introduction

This deliverable is designed to detail the management, processing, and protection of the scientific data collected, curated, and generated 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 research and development objectives shift, the plan can and will be refined. 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, Global Navigation Satellite System (GNSS), 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) and other pilots integral to the REALLOCATE project, inclusive of deployment, environmental, climatic, and device-specific data.

1.1 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 D1.3 – Data Management Plan represents the initial version of the DMP at M6, with subsequent updates scheduled for M18 (D1.4) and M36 (D1.5). As the project progresses, any data-related considerations should be viewed through the prism of this DMP.

The D1.3 is 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 D1.3 may not have a direct connection with any Key Exploitable Result, all data outcomes should be aligned with Intellectual Property Rights (IPR) guidelines.

1.2 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, relevance, 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.

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. Initially, the datasets will be presented with a broad overview, but as the project progresses, the details will become more specific. Section 7 showcases the 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. These data categories are broadly classified into the following:

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

- **Mobility Data:** This category includes information on how people move within urban areas. It covers origin-destination data, travel behavior 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 research 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, structured databases using SQL (Structured Query Language), 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.

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 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 outcome 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.

3.1 Data findability and provision for metadata

In line with the commitment outlined in the GA (Grant Agreement) and CA (Consortium Agreement), which define the management of results, access rights, and information non-disclosure, trusted open-access data repositories such as Zenodo, Open Data Repository or other tools (e.g., figshare, Dryad, Dataserve) will be utilised to ensure data findability. Also, as an OpenAIRE-compliant repository, Zenodo aligns well with the dissemination requirements. To provide comprehensive metadata, the Dublin Core and DataCite Metadata Schema are the preferred standard descriptive formats.

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 editors, while others will be assigned by the repositories/entities managing the repository, such as Zenodo. 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

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).

3.2 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/). 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.

A similar process will be applied to scientific publications. 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 is Google Drive under UCD control and 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 like Document edition-Content Management System (CMS), document management, calendar and tasks, contact lists. Authorization has been given to users one-by-one. It is designated to serve activities of document editing, archiving and file sharing among partners.

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.

3.3 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. Finally, the Data Hub will be built upon Kafka and Apache Spark data models, ensuring seamless interoperability. 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

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

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) **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

3.4 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). Data obtained will remain useable indefinitely.

By the time the DMP is delivered through D1.3 on M6 of the project, information about restrictions for the reusability of the project outcomes is still unknown. This information will be specified in the updated (D1.4) and 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. Subsequently, updates to the plan will be provided in 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 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].

5.1 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.

[2] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

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.

5.2 Repositories

The description of all the data to be shared will be placed in a research data repository (see Section 6). These repositories (being stored in secure servers owned by the Project Coordinator) will clearly point to all data entities shared that these can be accessed, mined, exploited, reproduced, etc.

These repositories have to sustain the data value and be safe in legal terms, as well as maintain for as long as possible all the stored data. Preferably, it should support analysis and track data usage, making use of a metadata system for facilitating search and discovery. Identification of the online repository that better suits the project's i.e. partners' needs and facilitate preservation, management and access to the scientific outputs of them is proposed as best practice. Therefore, three main repositories can be used:

- Zenodo (https://zenodo.org/) - is the repository recommended by the Open Access Infrastructure for Research in Europe (OpenAIRE).
- REALLOCATE website – Publications tab (https://reallocatemobility.eu/)
- IEEE dataports (https://ieee-dataport.org/...) – valuable and easily accessible data platform that supports research data management
- CKAN (https://ckan.org) – a powerful data management system that makes data accessible by providing tools to streamline publishing, sharing, finding, and using data. It is often used by governments, organizations, and institutions globally as an open-source data portal platform.

Further repository possibilities can be proposed and evaluated during the project.

For internal access to the datasets, among REALLOCATE partners, the datasets will be stored in the document repository of the project, Google Drive hosted 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.

Error! Reference source not found. illustrates the template for the dataset reporting is illustrated. 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	**Metadata**	Yes/No	Yes/No
Link where it can be accessed (if exists)			Adapted Format	eg. OpenAIRE

Figure 1. Data management monitoring template

7 Comprehensive Knowledge Archive (CKAN)

CKAN, which stands for Comprehensive Knowledge Archive Network, is a powerful open-source data management platform designed to cater to the varied needs of organizations and individuals focused on publishing, sharing, discovering, and utilising datasets. As such, CKAN emerges as an ideal platform that offers the essential infrastructure for efficient data management. Within REALLOCATE, its capabilities will be utilised to optimize data management efforts.

Recognized primarily by governments, research bodies, and data-focused entities, CKAN offers invaluable tools that cater specifically to publishing, sharing, and searching data. The platform's adaptability stands out, with its extensible and modular architecture that can be fine-tuned to resonate with the unique needs of different users. With API integration support, REALLOCATE can explore and develop customised applications, tools, or services that make optimal use of data stored within CKAN.

One of CKAN’s most distinctive features is its faceted search capability, which empowers users to quickly refine data searches based on specific criteria. Whether it's filtering by data type, organization, or format, users can swiftly pinpoint the exact datasets they need. Moreover, CKAN promotes transparency and clarity by allowing users to preview datasets through various visualization tools like maps, graphs, and tables before downloading or integrating them.

Another compelling feature is CKAN’s metadata-rich environment. Every dataset published is accompanied by detailed metadata, ensuring users comprehend the context, source, and relevance of the data. CKAN’s data retrieving toolset allows integration of datasets from multiple sources, ensuring that REALLOCATE can maintain an up-to-date repository without manual oversight.

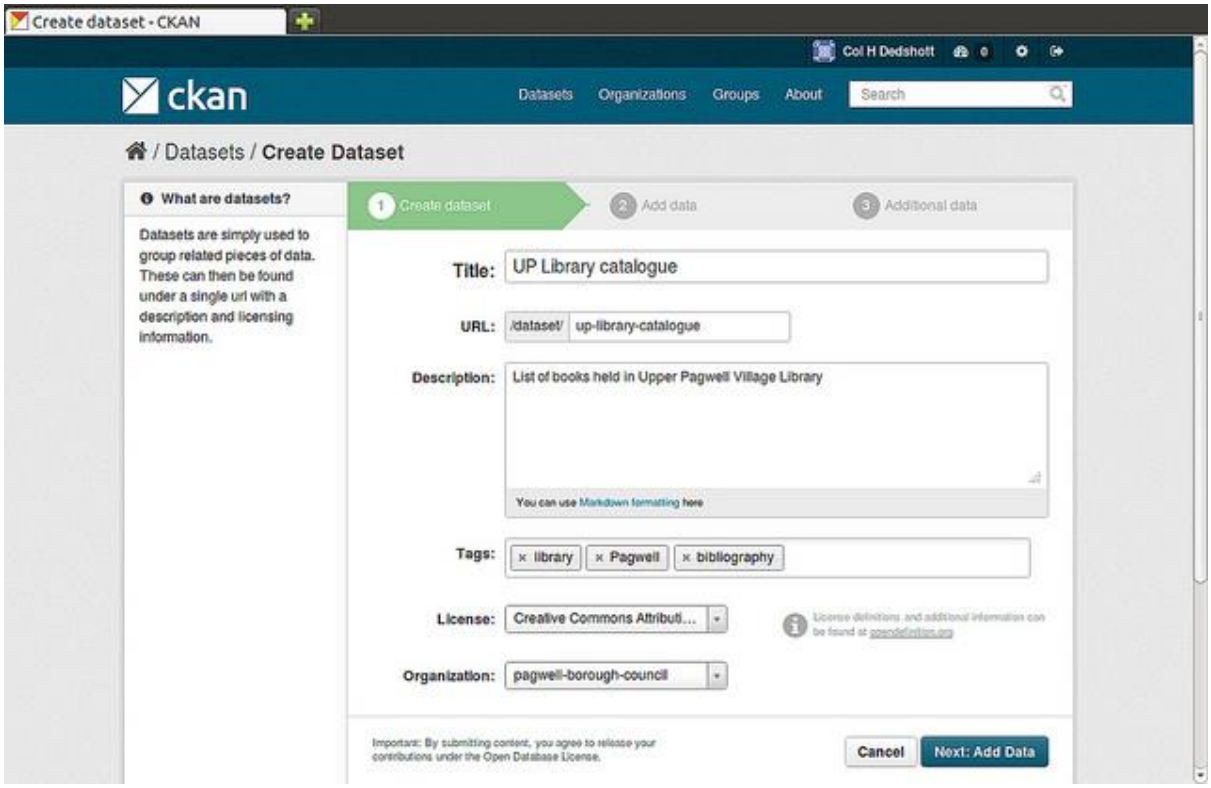

Figure 2. CKAN Interface Snapshot.

Additionally, to foster collaboration and openness, CKAN includes a commenting system. This facilitates discussions right at the dataset level, enabling users to share insights, raise queries, or provide feedback. Furthermore, its extensibility means that new functionalities can be added via plugins, ensuring it remains relevant for upcoming data challenges.

Furthermore, CKAN gives top priority to ensuring data security. With built-in features for granular access controls, administrators can easily define who can view, edit, or manage datasets. Roles and permissions can be assigned at both dataset and organization levels, ensuring that sensitive data is adequately safeguarded. SSL support ensures data transmission remains encrypted, adding another layer of security during data access and upload.

CKAN's design also embraces interoperability. Its RESTful API ensures seamless integration with a multitude of external systems, tools, and platforms. For projects like REALLOCATE, this means the ability to harmonize data operations across diverse IT ecosystems, making data exchange and synchronization seamless.

Finally, CKAN has a dynamic global community of developers, data enthusiasts, and institutional users, who collectively enhance the platform, contribute extensions, and provide support, enabling to effectively filling any feature gap.

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, the access is enabled to the user.

The non-sensitive data will be stored on the Data Hub, which will be protected by the State-of-the-Art security systems (ransomware protection). It will run at secure premises behind DMZ (Demilitarized zone) and firewall control, besides internal and additional security mechanisms as SSL (Secure Sockets Layer) server implementation, authorization and access control. It additionally will employ an enterprise level backup plan implemented in the event of system failure (i.e., daily delta changes back-up, and weekly full back up on the entire content of the repository).

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.

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 issue

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[3], 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 will be 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 related to mobility and environmental data from various sources. 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.

[3] Reliability, honesty, respect, accountability – European Code of Conduct for Research Integrity

- D5.4: This deliverable involves evaluating the ethical implications of the interventions and their impacts on various dimensions, including safety, environment, social inclusiveness, and transformative governance. This evaluation indirectly relates to ethical principles. Furthermore, the assessment process may involve the collection and analysis of data, which could include personal data. Therefore, it's vital to ensure that data handling aligns with data protection regulations, and measures are in place to safeguard personal data throughout the assessment. Upholding research integrity is also a key aspect, ensuring that assessments are conducted transparently and impartially, contributing to the ethical foundation of the project's activities.

11 Future Work and conclusions

The initial version (D1.3) of the DMP aims to provide a broad overview of FAIR data management, aligning with the EC's goals for data sharing and knowledge dissemination. As the project advances, the DMP will evolve to become more detailed and specific. This evolution may be driven by factors such as the identification of new datasets, changes in Consortium policies, the emergence of Open Call projects, or modifications to the project team. Therefore, the content of the DMP will remain dynamic, becoming progressively more specific as the project moves towards its milestones at M18 (D1.4) and M36 (D1.5).

A. Appendix

1. 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	☐

	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.	☐