

D1.6 Data Management Plan – DMP (V3 Final)

01/02/2024 Author(s): Renee OBREGON GONZALEZ (AKKODIS)



MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).

2

Summary sheet

Deliverable Number	D1.6
Deliverable Name	D1.6 - Data Management Plan – DMP (V3 Final)
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Renee OBREGON GONZALEZ (AKKODIS)
Contributing Partner(s)	AKKODIS
Peer Review	HERE, ICOOR, POLIS, URV
Contractual Delivery Date	31-01-2024
Actual Delivery Date	30-01-2024
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	49
WP/Task related to the deliverable	WP1/T1.6
WP/Task responsible	AKKODIS
Document ID	MobiDataLab-D1.6-DataManagementPlanV3Final_v1.0.docx
Abstract	This document defines all the procedures to handle the data collected or generated and how they are processed and preserved in the MobiDataLab project.

Legal Disclaimer

MOBIDATALAB (Grant Agreement No 101006879) is a Research and Innovation Actions project funded by the EU Framework Programme for Research and Innovation Horizon 2020. This document contains information on MOBIDATALAB core activities, findings, and outcomes. The content of this publication is the sole responsibility of the MOBIDATALAB consortium and cannot be considered to reflect the views of the European Commission.







Project partners

Organization	Country	Abbreviation
AKKODIS	France	AKKODIS
HERE GLOBAL B.V.	Netherlands	HERE
CONSORZIO INTERUNIVERSITARIO PER L'OTTIMIZZAZIONE E LA RICERCA OPERATIVA	Italy	ICOOR
POLIS - PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES	Belgium	POLIS
UNIVERSITAT ROVIRA I VIRGILI	Spain	URV

Document history

Version	Date	Organization	Main area of changes	Comments
0.1	22/12/2023	AKKODIS	All	Table of Content
0.2	08/01/2024	AKKODIS	All	Draft version
0.3	17/01/2024	AKKODIS	All	Contribution
0.4	17/01/2024	ICOOR	All	Peer Review
0.5	17/01/2024	POLIS	All	Peer Review and feedback
0.6	18/01/2024	AKKODIS	All	Rework
0.7	22/01/2024	URV	Section 1 and 4	Peer Review
0.8	24/01/2024	HERE	Table of contents	Peer Review
0.9	30/01/2024	AKKODIS	All	TL + Coordinator Quality Check
1.0	30/01/2024	AKKODIS	All	Final Version and Submission





4

Executive Summary

This report is the third deliverable of Task 1.4 "Data Management" and describes the updated Data Management Plan (DMP) for the MobiDataLab project, funded by the EU's Horizon 2020 Programme under Grant Agreement n°101006879.

The purpose of the DMP is to provide an overview of all datasets collected and generated by the project and to define the MobiDataLab consortium's data management policy that is used regarding these datasets.

The first DMP (deliverable D1.4) followed the structure of the Horizon 2020 DMP template and reported on the datasets used and produced by the project. It also defined the general policy and approach to data management in MobiDataLab that handles data management-related issues on the administrative and technical levels. This included for example topics like data and metadata collection, publication and deposition of open data, and the data repository infrastructure.

The second DMP (deliverable D1.5) was implemented as a "living" DMP that was continuously updated throughout the course of the project. It reflects the status of the collected, processed or generated data and it followed methodology and standards, how this data was going to be shared and/or made open, and how it will be curated and preserved.

The third DMP (this deliverable D1.6) is an update of the previous version and will be implemented as the end of the project approaches, but it will provide an open window to new opportunities and developments. This clearly must comply with the FAIR principles to be able to provide an opening for a possible continuation.





5

Table of contents

1.	INTRODUCTION	9
	1.1. PROJECT OVERVIEW	9
	1.2. PURPOSE OF THE DELIVERABLE	9
	1.3. INTENDED AUDIENCE	10
	1.4. STRUCTURE OF THE DELIVERABLE AND ITS RELATIONSHIP WITH OTHER WO	RK
	PACKAGES/DELIVERABLES	10
2.	MOBIDATALAB DATA	12
	2.1. PURPOSE OF THE DATA COLLECTION AND UTILITY	12
	2.2. DATA CATEGORIES	12
	2.2.1. Internal administrative data	13
	2.2.2. Open research data	13
	2.2.3. Evaluation data	14
	2.2.4. Technical data	16
	2.3. DATASET DESCRIPTION	18
3.	DATA MANAGEMENT IN MOBIDATALAB	21
	3.1. ABOUT DATA MANAGEMENT IN MOBIDATALAB	21
	3.2. ROLES AND RESPONSIBILITIES	22
	3.3. DATA MANAGEMENT PLAN EDITION	23
	3.4. TRANSPORT CLOUD AND OPEN KNOWLEDGE BASE	24
	3.5. OPEN DATA USE	25
	3.6. NON-OPEN DATA USED	26
	3.7. OPEN DATA CREATED	26
	3.8. "NON-OPEN" DATA CREATED	26
4.	FAIR DATA PRINCIPLE	27
	4.1. MAKING MOBIDATALAB (META)DATA FINDABLE	
	4.2. MAKING DATA OPENLY ACCESSIBLE	30
	4.3. MAKING DATA INTEROPERABLE	33
	4.4. INCREASE DATA RE-USE	
5.	GDPR	
	5.1. PERSONAL DATA COLLECTION, PROCESSING, AND RE-USE	
	5.2. COMPLIANCE WITH GDPR RECITAL 78	
	5.3. DATA PROTECTION AGENCY NOTIFICATION	
	5.4. COMPLIANCE WITH ARTICLE 49 OF THE GDPR	
	5.5. SAFE HARBOUR AND EU-US PRIVACY SHIELD CONSIDERATIONS	
	5.6. ACTIVITIES DEDICATED TO ETHICAL CONSIDERATIONS	
	5.7. MINIMUM RESORT TO EXCEPTIONS AND DEROGATIONS	
	5.8. ACTIVITIES DEDICATED TO THE PROTECTION AND SECURING OF PERSONAL	. DATA 40
	5.9 SHARED INFORMATION AND PERSONAL DATA	
6	ALLOCATION OF RESOURCES	42
7	DATA SECURITY	44
	7.1.1. Storage	
	7.1.2 Authentication and authorisation	11
	I. I.∠. ∩uuisiiuuauuii aiiu auuiviisauvii	





Funded by the European Union

	7.1.3. Data protection	45
8.	CONCLUSION	46
9.	REFERENCES	47

List of figures

Figure 1: MobiDataLab Zenodo Repository	21
Figure 2: MobiDataLab Project in OpenAIRE	22
Figure 3: MobiDataLab page on DMPOnline	24
Figure 4: The relationship between FAIR and Open data	28
Figure 5: Open access to research data	31

List of tables

Table 1:	MobiDataLab categories	12
Table 2:	Mobility data categories related to use cases.	18
Table 4:	Dataset description template	19
Table 5:	Dataset description in the MobiDataLab catalogues	20





Abbreviations and acronyms

Abbreviation	Meaning
API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
D	Deliverable
DCAT	Data Catalogue Vocabulary
DMP	Data Management Plan
EC	European Commission
EOSC	European Open Science Cloud
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GeoJSON	Geographical features based on JSON
GDPR	General Data Protection Regulation
GTFS	General Transit Feed Specification
H2020	Horizon 2020
HTTPS	Hypertext Transfer Protocol Secure
IPR	Intellectual Property Rights
IPsec	Internet Protocol Security
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MaaS	Mobility as a Service
NeTEx	Network timetable Exchange



OA	Open Access
ОКВ	Open Knowledge Base
ORD Pilot	Open Research Data Pilot
OSM	Open Street Map
REST	Representational State Transfer
RDF	Resource Description Framework
SaaS	Software as a Service
SPARQL	SPARQL Protocol and RDF Query Language
URV	Universitat Rovira I Virgili
V	Version
W3C	World Wide Web Consortium
WP	Work Package





1. Introduction

1.1. Project overview

There has been an explosion of mobility services and data sharing in recent years. Building on this, the EU-funded MobiDataLab project works to foster the sharing of data amongst transport authorities, operators and other mobility stakeholders in Europe. It develops knowledge and a cloud solution aimed at easing the sharing of data. Specifically, the project is based on a continuous co-development of knowledge and technical solutions. It collects and analyses the advice and recommendations of experts and supporting cities, regions, clusters and associations. These actions are assisted by the incremental construction of a cross-thematic knowledge base and a cloud-based service platform, which will improve access and usage of data-sharing resources.

1.2. Purpose of the deliverable

The Data Management Plan (DMP) defines all the procedures to handle the data collected or generated within the project, and how they are processed and preserved. It describes the approach to making the MobiDataLab data Findable, Accessible, Interoperable and Reusable (FAIR1¹) by indicating what data will be generated, collected and processed, what standards will be applied, how research data will be preserved and what parts of the datasets will be shared for evaluation purposes and to comply with Open Research Data Pilot (ORD Pilot) requirements. The document also addresses ethical and confidentiality issues and some data security principles. This deliverable is the last version of this document as the project comes to its ending phase.

- The initial version D1.4 (M6, i.e., July 2021) has described the Data Management Plan according to the current evolution of the project. It identifies an initial set of data categories that will be involved in the project and proposes the data management process that will be followed in future developments.
- The updated version D1.5 (M18, i.e., July 2022) has provided an update of the Data Management Plan including the description of the datasets and the potential evolution of the procedures defined at the beginning of the project.
- The final version D1.6 (M36, i.e., January 2024) includes the final description of the datasets (the mobility data used in the project), and procedures followed.

¹ OpenAIRE – <u>How to Make Your Data FAIR?</u>





1.3. Intended audience

The dissemination level of this D1.6 deliverable is public, and it is available to the members of the consortium, the European Commission (EC) services and those external to the project.

This document is primarily intended to serve as an internal guideline and reference for all of the partners of the consortium, reference group and advisory board members. This deliverable is also of high interest to anyone willing to understand the data management procedures followed during the project as part of the process of reusing the published datasets.

1.4. Structure of the deliverable and its relationship with other work packages/deliverables

This deliverable serves as an entry point to understand the project-wide approach to data management in MobiDataLab. It provides an overview of data management at the project level with a stronger focus on research data management, as required by the H2020 ORD Pilot programme. This version will also cover categories of mobility data that have been used during the project.

This deliverable is structured as follows:

- Section 1 is an introduction to the project and the deliverable.
- Section 2 covers the types of data collected and managed by MobiDataLab, including mobility data categories.
- Section 3 focuses on the data management plan of MobiDataLab, more particularly on how the diverse types of data will be handled and curated (what type of tools and repositories are used and who are the responsible members).
- Section 4 provides information on how MobiDataLab is making data more findable, accessible, interoperable, and reusable by targeting each FAIR principle. It provides examples of how data is collected, shared, processed, generated, and made accessible for verification and reuse.
- Section 5 goes on how the project will comply with the General Data Protection Regulation (GDPR).
- Section 6 provides information on the allocation of resources within the project.
- Section 7 covers the plan for data security, including data storage and authentication.
- Section 8 provides a short conclusion of the deliverable.

This deliverable is linked to other deliverables; in particular, deliverables from WP2 (Open Knowledge Base) and WP3 (New data sharing services and business models) as an important part of the research work that has been carried out in these work packages.

The Data Management Plan also serves as a reference document for the implementation of the MobiDataLab Transport Cloud prototype (WP4), whose aim was to make the mobility data FAIR through our prototype platform.





In particular, the "Reference Data Catalogue" (Task 4.2 / D4.3 and D4.4) refers to the Findable dimension, the "Data Access Services" (Task 4.3 / D4.5 and D4.6) refers to the Accessible dimension, whereas the "Data Processors" (Task 4.4) and "Data Protection Tools" (Task 4.5) refer to the Interoperable and Reusable dimensions respectively.





2. MobiDataLab data

MobiDataLab handles different types of data which can be organized into four categories: administrative data, open research data, evaluation data and technical data. This section will describe these data categories while providing information on their management and sharing.

2.1. Purpose of the data collection and utility

The objective of the data collection is to produce knowledge, develop cloud solutions, and propose a methodology and tools to ease data sharing and participate in the establishment of a data-sharing culture in Europe. The data collection also aims to assess the current market potential and the project's new data sharing-services and business models. More specifically, the data collection in MobiDataLab can be linked to three main pillars: the Open Knowledge Base, the Transport Cloud and the Labs (Living and Virtual).

2.2. Data categories

This section focuses on the categories of data that are handled during the project, which are: administrative data, technical data, evaluation data, and open research data. A short description of these categories is given in Table 1 while more details on the corresponding data management are provided in the upcoming subsections.

Table 1: MobiDataLab categories

Category	Short description
Internal administrative data	This category refers to the data generated/shared internally for administrative and management activities for historical purposes and follow-ups within WP1 such as meeting minutes, recordings, internal reports and others. It is discussed in more detail in 0.
Open research data	This category contains the data and results that will be published by the project to comply with ORD Pilot requirements. The open research data is a subset of the evaluation data which is itself a subset of the technical data. It is addressed in 2.2.2.
Evaluation data	The evaluation data concerns the data used to compute the KPIs measured for data sharing culture. It includes the technical data concerning the quantitative and qualitative evaluations of the Transport Cloud and other related business models for data-sharing services. This category is discussed in more detail in 2.2.3.
Technical data	This category includes the data related to the technical developments of WP4, WP5 and solutions identified in WP2. This category is discussed in more detail in 2.2.4. Mobility Data Sharing is a sub-category of this section.





2.2.1. Internal administrative data

This category refers, on one hand, to the data produced by the project management activities for historical purposes and follow-ups within WP1 such as meeting minutes, recordings, internal reports and others.

In terms of deliverables, relevant linked information is presented in D1.1 "Collaborative portal", D1.2 and D1.3 "Project, Innovation and Quality management plan". The data is collected by the management team including the project manager, the Work Package (WP) leaders and task leaders. These deliverables are stored in Microsoft Teams, which is a collaborative application providing different basic functionalities and extensions. Preliminary guest accounts have been requested for all the members of the consortium to access all available functionalities linked to the project. This administrative data is confidential and is available only to the members of the consortium.

On the other hand, administrative data also refers to another type of data linked to the project management, including:

- Data management:
 - Information included in D1.4, D1.5 and D1.6 "Data Management Plan".
- Data exploitation:
 - o Information included in D6.8 and D6.9 "Exploitation Plan".
- Project activities communication:
 - Information included in D6.1 "Dissemination report".
 - o Information included in D6.2, D6.3 and D6.4 "Reporting on MobiDataLab events".
- Activities concerning project stakeholders (specifically, there are deliverables):
 - o Information included in D1.7 "Report on Expert Committee activities".
 - Information included in D6.5 "Stakeholder group activities report".
 - o Information included in D6.6 and D6.7 "Project cooperation activities report".

This data is open to the public and is accessible via the project website: <u>MobiDataLab</u>, as well as on CORDIS² (where it is possible to find the latest version).

2.2.2. Open research data

Open research data is made available to the research community as part of the state of the art, to encourage comparison between methodologies with other approaches or projects. Some of the non-confidential KPIs published at the end of the project could be used to evaluate in a cross-Living Lab context or a comparison with future approaches regarding the integration of MobiDataLab's technologies.

² <u>https://cordis.europa.eu/project/id/101006879</u>





The open research data category refers to the data resulting from the research work done. This category includes findings presented in most of the deliverables from WP2 and WP3.

WP2 has consolidated an Open Knowledge Base³ derived from the most important projects and initiatives implemented to date in the domain of transport data sharing. It has also identified new requirements from predefined use cases and a list of concrete problems faced by mobility stakeholders.

The output deliverables are:

- D2.1 "Legal and Regulatory data sharing gap analysis",
- D2.2 "Recommendations on data sharing legal frameworks",
- D2.3 "State of the art on Mobility and Transport data protection technologies",
- D2.4 "State of the art on Mobility data sharing standards",
- D2.5 "Report on new Mobility data sharing standards",
- D2.6 "Report on enabling technologies for Transport Cloud",
- D2.7 "Data Governance assessment",
- D2.8 "Data Governance recommendations" and,
- D2.9, D2.10 "Use cases definition v1 and v2".

The overall WP3 goal was to analyse the market's state-of-the-art and enhance the potential impact of digitalization and data sharing on different actors and different areas of mobility and transport (economic, social, environmental, etc.). The following reports were produced as deliverables:

- D3.1 "Actors' needs and cooperation framework report",
- D3.2 "Data sharing market technological developments monitoring report",
- D3.3 "Market Gap Analysis report",
- D3.4 "Data sharing business and revenue models" and,
- D3.5 "Societal and Environmental Impacts of data sharing assessment framework".

These deliverables are open to the public and are accessible via the repository chosen by the consortium and on the website of the project.

2.2.3. Evaluation data

The evaluation activity requires the collection of a subset of the data generated during the technical development phase. The publication of evaluation data consists of the selection and publication of a subset of the data used as research data at the end of the project following the ORD Pilot requirements (to learn more about it, please consult the D1.5 section 6.1.1).

Only the data that the partners agree to make openly available will be published and will constitute open research data.

³ <u>https://mobidatalab.eu/knowledge-base/about/</u>





The evaluation data category includes the technical data concerning the quantitative and qualitative evaluations of the Transport Cloud and other related business models for data-sharing services. This includes the Key Performance Indicators (KPI) measured for the data-sharing culture, to determine a list of requirements to promote data sharing in the context of exploration and collaboration among actors, and to draw a projection of the data-sharing ecosystem in transport and mobility for the coming years in terms of qualitative estimations.

WP3 aimed to guide the evaluation of the new data-sharing services and business models. In particular, the objective of Task 3.5 "Societal and Environmental Impacts of Data Sharing Assessment Framework" was to define an evaluation methodology to assess the impact of data-sharing services provided by the new Transport Cloud and the related business models on society and the environment. The task proposed a methodology to assess the relevant KPIs, by considering both quantitative and qualitative evaluations. A specific focus was dedicated to creating a survey that was used in Task 5.2 "Quantification and measurement of the data exchange culture" to evaluate:

- the acceptance of the Transport Cloud,
- the business models' acceptance.

By month 21 of the project, this evaluation framework was ready and validated by the General Assembly in March 2023 for further use in the Living Labs instances. The deliverables of Task 5.2 are:

- D5.2 "Report on Quantification and measurement of the data exchange culture"
- D5.3 "Analysis and conclusions on the data exchange culture"

This task of WP5 measured the impact of the data exchange culture by assessing the impact of Transport Cloud services on use case challenges. It also assessed the users' acceptance of the Transport Cloud and the business models. To that end, it used the framework developed in T3.5, the assessment realised, qualitatively but also quantitively when possible, the hindrances and opportunities to data sharing raised by participants of Living Labs and experts, and how those are satisfied with data sharing initiatives such as the Transport Cloud. The survey regarding data sharing culture evaluated other use cases as well, exploring and expanding on the recognized opportunities paving the way for new projects, creating requirements for further research and providing input for improving the Transport Cloud's functionalities. The first output of Task 5.2 was an analysis of the impact of the Data Exchange Culture known as the D5.2 "Report on Quantification and Measurement of the Data Exchange Culture". The upcoming result will be produced by the end of M36 as the D5.3 "Analysis and conclusions on the data exchange culture". This considers feedback from the participants of the Living Lab, who used the Transport Cloud services during the project, and who after the project.

This evaluation data is accessible to the public via the repository of the consortium⁴ and on the website of the project.

⁴ <u>https://Zenodo.org/communities/mobidatalab</u>





2.2.4. Technical data

Technical data is required when performing the evaluation activities. The technical data and the evaluation data are then useful to the evaluation task, to calculate and produce the KPIs necessary to evaluate the project approach.

The technical data category includes the data related to the technical developments of the MobiDataLab project, which corresponds to the contents of WP4 and WP5. WP4 leveraged the open solutions identified in WP2 to prototype a platform for searching, accessing and fusing multimodal mobility data in the cloud. WP4 aimed to showcase the most effective tools to facilitate access to and exchange of mobility data (by humans and machines) for Mobility as a Service (MaaS) companies and developers. The resulting deliverables were:

- D4.1, D4.2 "Transport Cloud Architecture dossier v1 and v2".
- D4.3, D4.4 "Reference Data Catalogue v1 and v2".
- D4.5 "Data Access services v1 Pilot set of data provided via services and initial interfaces".
- D4.6 "Data Access services v2 Description of data access interfaces".
- D4.7, D4.8 "Data Enrichment Processors v1 and v2".
- D4.9, D4.10 "Data protection tools v1 and v2".

In WP5, the project evolved into concrete and real-time data usage fostered by the Living Labs implementation in different locations and through the Virtual Lab that supported the process and fostered data exchange during the project. In MobiDataLab, the Living Labs serve two main purposes. Firstly, to assess and measure, qualitatively and quantitatively, the data sharing culture. Secondly, to generate new tools, technologies and insights through the utilization of the MobiDataLab tools and data, leading to an evaluation of the MobiDataLab tools and technologies.

The deliverables include:

- **D5.1** "The Virtual Lab".
- **D5.4** "Living Labs execution plan".
- **D5.5** "Report on Living Labs monitoring".
- D5.6 "Report on #datathon".
- D5.7 "Report on #hackathon".
- D5.8 "Report on #codagon".

The deliverables mentioned above are public except for D5.4, D5.6, D5.7 and D5.8 which are confidential and are available only for the members of the consortium. The partners have agreed on the shareable and confidential data. Nonetheless, deliverable D5.5 will provide a summary of the most important aspects covered in these deliverables without involving any confidential content.

Concerning to the challenges defined in WP2 in deliverables 2.9 and 2.10 and proposed through the Labs of WP5, datasets were collected through the Reference Data Catalogue.





To facilitate the discovery of data, the use case data combined transport data and other kinds of data (since data discoverability should not depend on its domain and application); for example, accessibility data, demographic and socioeconomic data, geographical and environmental data, tourism data and land use data. The Reference Data Catalogue's datasets were organized as groups to be more easily manageable, searchable and findable. A dataset could belong to several groups, but a dataset had a unique identifier. The groups were defined in relation to use case challenges of the MobiDataLab Labs, but also by the identified transportation themes and categories of the work done by DATA4PT⁵ and the INSPIRE Directive (Directive 2007/2/EC)⁶.

Mobility data sharing categories

Mobility data is continuously produced, at a higher pace and in various forms. The collection and aggregation of new sources of data from various producers has become a necessity, with data-sharing principles at its heart. The main categories of datasets collected concerning the use cases were:

- Data for Estimated Time of Arrival (ETA) computation (traffic real-time and historical data, static map data, weather, rest time regulations, planned events like road closures etc.)
- Operational data (telematics data of vehicles, location of vehicles, completed stops, tour
- plans, driver shift time).
- Public transport data (static data, transportation lines, schedules, stop points, stop areas, realtime/dynamic data, disruptions, traffic alerts, next arrivals and departures, vehicle occupancy, etc.)
- Other transport data (free-floating, ridesharing, road traffic)

GPS tracks of mobile phone users⁷, origin-destination⁸ and daily trajectory logs⁹ are all examples of privileged¹⁰ datasets that are accessible in the data catalogues related to journey planning and which are also examples of anonymized trajectories of personal traveller data.

¹⁰ They were provided in the context of the Labs. They are normally private datasets but can be openly shared in the context of MobiDataLab Labs.





⁵ <u>https://data4pt-project.eu/siri-webinar-material/</u>

⁶ https://knowledge-base.inspire.ec.europa.eu/legislation/inspire-directive_en

⁷ https://ckan.mobidatalab.eu/dataset/gps_raw_data_france

⁸ https://ckan.mobidatalab.eu/dataset/od_data_leuven

⁹ https://ckan.mobidatalab.eu/dataset/trajectory-logs-datasets

Table 2: Mobility data	a categories	related to	use cases.
------------------------	--------------	------------	------------

Mobility data category	Short description
Journey Planning	Datasets about travel planning. Also includes information about Software as a Service (SaaS).
Freight	Goods of cargo, cargo control data, cargo network, cargo places, real-time-freight data and scheduled-freight.
Micro-mobility	Datasets related to micro-mobility services or features.
Mobility & Transport	Information about public transport, transport networks, roads, transport infrastructure and all kinds of information related to mobility and transport.
Parking	General parking information. Information about parking fees for a right or privilege to use a particular road.
Shared mobility	Datasets of shared transport including areas, cost and offers. Shared transport systems include carsharing, bicycle sharing, carpooling and vanpools (aka ridesharing or lift-sharing), real-time ridesharing, slugging (casual carpooling), community buses and vans, demand responsive transit, paratransit, a range of taxi projects and even hitchhiking and its variants.
Ticketing data	Ticketing gives access to services. This is done through several functions, therefore here you will find datasets related to information, booking, fares, physical access, payment and control.
Traffic	Road traffic control, road traffic or sea traffic.
Transport Networks	Road, rail, air and water transport networks and related infrastructure. Includes links between different networks. Also includes the trans-European transport network as defined in Decision No 1692/96/EC of the European Parliament and of the Council of 23 July 1996 on Community Guidelines for the development of the trans-European transport network (1) and future revisions of that Decision.

2.3. Dataset description

This section provides guidelines on how to describe the different types of datasets collected and shared by MobiDataLab after the end of the project concerning ORD Pilot which aims to improve and maximize access to, and re-use of research data generated by Horizon 2020 projects.

The description of the different datasets should provide information on their reference, file format, standards, methodologies, metadata, and repository to be used. The following table is a template that can be used to describe the datasets of the project. This template follows the Data on the Web Best Practices, recommended by the World Wide Web Consortium (W3C)¹¹.

¹¹ <u>https://www.w3.org/TR/dwbp/</u>





Identifier of the dataset (URI)	Identifies each dataset by a persistent URI ¹² .
Name of the dataset	Name of the dataset
Descriptive metadata	Descriptive metadata can include the following overall features of a dataset (keywords, date of publication, spatial coverage, date of last modification, themes/categories covered by a dataset). Descriptive metadata can also include the overall features of the different distributions of the dataset (distribution name, description, media type, etc.)
Data provenance	Provides complete information about the origins of the data (e.g., the publisher, the contact point, etc.) and any changes made after publication. Data provenance can be provided using an ontology recommended to describe provenance information, such as W3C's Provenance Ontology.
Structural metadata	Provides metadata that describes the schema and the internal structure of distribution, usually the properties or columns of the dataset schema.
Data licenses	A link to, or embedded copy of, a human-readable license agreement and/or machine-readable license agreement (e.g., ODBL, Open Data, etc.).
Data formats and standards	Machine-readable standardised data format under which the data is made available. Standardised data formats enable interoperability as well as future uses.
Personal data (GDPR)	Provides information about personal data and mentions if the data is anonymised or not. Tell if the dataset entails personal data and how this issue is taken into account.
Data preservation	The preservation guarantees and the data storage during and after the project (for example databases, institutional repositories, public repositories, etc.).

Table 3: Dataset description template

In the case of our data catalogues, the metadata of the dataset descriptions changes, but comprises most of the elements in the template of the W3C. It is important to mention that in GeoNetwork¹³ metadata can be adapted to various profiles; therefore, the table examples are not exhaustive.

Humans should be able to interpret the nature of the dataset and its distributions, and software agents should be able to automatically discover datasets and distributions. The machine-readable version of the descriptive metadata is provided using the vocabulary recommended by W3C to describe datasets, i.e., the Data Catalog Vocabulary (DCAT)¹⁴.

An example description of a mobility-related dataset is available in the "Data on the Web Best Practices" published by the W3C¹⁵.

¹⁵ <u>https://www.w3.org/TR/dwbp/dwbp-example.html#dataset-description</u>





¹² <u>http://philarcher.org/diary/2013/uripersistence/</u>

¹³ <u>https://geonetwork.mobidatalab.eu/GeoNetwork-4.0.5-0/srv/eng/catalog.search#/home</u>

¹⁴ https://www.w3.org/TR/vocab-dcat-3/

CKAN	Dataset description	GeoNetwork
Title	Name of the dataset	Title
Unique identifier	Identifier of the dataset (URI)	Resource Identifier
Organization	Data provenance	
Description	Descriptive metadata	
Visibility	Descriptive metadata	
Licence	Data licenses	Legal constraints
Tags	Descriptive metadata	Keywords
Source	Data provenance	
Groups	Descriptive metadata	INSPIRE themes
State	Descriptive metadata	Status
Created	Descriptive metadata	
Extra fields: Language	Structural metadata	Language
Author	Data provenance	
Author Email	Data provenance	Contact for the resource
Maintainer	Data provenance	
Maintainer Email	Data provenance	
Extra fields: Frequency	Descriptive metadata	Update frequency
Last Updated	Descriptive metadata	Update
Multiple formats	Data formats and standards	Format
Extra fields	Descriptive metadata	Metadata language
Version	Descriptive metadata	

Table 4: Dataset description in the MobiDataLab catalogues





3. Data management in MobiDataLab

3.1. About data management in MobiDataLab

MobiDataLab as an H2020 project is obliged to provide frequent updates of its Data Management Plan (DMP) that describes what data the project will use and produce, whether and how data produced will be exploited or made (openly) accessible for verification and re-use, and how these data will be curated and preserved after the end of the project. Open data must be put into a public (research data) repository. MobiDataLab chose Zenodo (Figure 1). The repository has to be OpenAIRE¹⁶ compliant to enable the harvesting of metadata.



Figure 1: MobiDataLab Zenodo Repository

¹⁶ OpenAIRE is a socio-technical infrastructure for Open Scholarly Communication in Europe.





Funded by the European Union A project-wide data management policy that handles these issues on the administrative and technical levels have been established and described in MobiDataLab deliverable D1.4 Data Management Plan v1. This includes for example topics like data and meta-data collection, personal data treatment, the data repository infrastructure and the mandatory compliance to <u>OpenAIRE</u> - the Open Access Infrastructure for Research in Europe (Figure 2).

	Search Deposit Link Data sources Funders	Q, Sign in
]
MobiDataLab Labs for prototyping future Mobility Data sharing cloud Open Access Mandate for Publications and Research dat	d solutions ta • 🖸 <u>Project</u> • 01 Feb 2021 (Started) - 31 Jan 2024 (Ending) • On going	 Oversity Views Downloads 121
Funder: European Commission Project code: 1010	06879 Call for proposal: H2020-MG-2020-SingleStage-INEA	
Funded under: H2020 RIA Overall Budget: 2,995,5	560 EUR Funder Contribution: 2,995,560 EUR	
Summary Publications (47) Other research pro	oducts (1) DMPs (4) Metrics	ıl, Statistics
Description The overall objective of MobiDataLab is to authorities, operators, industry, innovators) a development of a data sharing culture in Eur development of knowledge and of technical producers and consumers in the transport a problem-solving oriented Labs, the collectio and supporting cities/regions/clusters/assoc thematic knowledge base and of a cloud-bas data sharing resources. MobiDataLab leverag to: - present mobility data providers with rec	propose to the mobility stakeholders (transport organising replicable methodology and sustainable tools that foster the ope and beyond. MobiDataLab is based on a continuous co- solutions for data sharing with high involvement of all data and mobility landscape. This will be put in action through on and analysis of advice and recommendations of experts bations aided by the incremental construction of a cross- ied service platform, which will federate access and usage of ges on the legal, technological and economic opportunities commendations on how to improve the quality, accessibili	Partners HERE GLOBAL B.V., HOVE, CNR, AETHON ENGINEERING, URV, ICOOR, KUL, AKKA I&S, F6S IE, POLIS
C Powered by the OpenAIRE Graph	Last update of records in OpenAIRE: Nov 11, 2023	Found an issue? Give us feedback

Figure 2: MobiDataLab Project in OpenAIRE

3.2. Roles and responsibilities

Data management activities concern the whole project and need to be coordinated and monitored both at project and work package levels. Data management is also linked to the publication of project results and thus dissemination activities. Therefore, the following roles and responsibilities can be identified:





- The Data Protection Officer (T1.4 task leader) is responsible for:
 - Developing the data management plan and policy in cooperation with the project management in WP1 and the technical partners.
 - Developing a user guide for the usage of MobiDataLab living DMP.
 - Advise data controllers and processors within the MobiDataLab project on the processing of personal data, training of researchers and assignment of responsibilities.
 - Assist in risk assessment of personal data processing.
 - Monitoring data management activities (both collection and publication) and deadlines and sending reminders to WP data managers.
 - Providing support to WP data managers.
 - Coordinating the writing of the DMP deliverable documents (D1.4-5-6).
 - Providing solutions for specific issues in accordance with the project management.
 - Cooperate with any national or European supervisory authority and function as the contact point for the project with such authorities.
- The Work Package Data Managers are responsible for:
 - The implementation of the data management policy in their respective WPs.
 - Monitoring data management activities and deadlines and sending reminders to partners.
 - Offering customized help and further guidance, to the respective WP task I WP task leaders and participants, for using the living DMP of MobiDataLab.
 - Asking partners for missing information or clarifications.
 - Providing input to the DMP deliverable documents (D1.4-5-6) by analysing and summarising the WP- specific datasets listed in the living DMP of MobiDataLab.
 - Offering customized help and further guidance for publishing open data and open-source software.
 - Monitoring that those open results (data and software) are deposited in the default repository or a complementary OpenAIRE-compliant repository and sending reminders to partners.
 - Monitoring those open results available in OpenAIRE are properly linked (https://www.openaire.eu/participate/claim) with MobiDataLab.
 - Contacting the quality assurance and ethics committee in case of questions and ethical and privacy issues that may forbid the publication of the data.
 - Ensuring that the metadata of data used and produced at the Work Package level is made available in the living DMP according to the MobiDataLab data management policy and guidelines promptly.

AKKODIS, represented by Thierry Chevallier as Coordinator of MobiDataLab, acts as the point of contact for Data Protection Issues and is accompanied by Renee Obregon acting as Data Protection Officer (DPO) in the project.

WP Data Managers are named inside each Work Package.

3.3. Data management plan edition

DMPonline is a tool intended to help in the (co)drafting, revision and sharing of DMPs. It allows one to write a plan collaboratively and to request proofreading directly from the application. MobiDataLab uses DMPOnline to gather more efficiently inputs from partners and WP leaders.





nomail@dcc.ac.uk ssion	
MobiDataLab, Labs for prototyping fu	Iture Mobility Data sharing cloud solutions
Project Details Contributors Plan overview Initial DMP De	etailed DMP Final review DMP Share Download
* Project title MobiDataLab. Labs for prototyping future Mobility Data sharing	Select Guidance
 mock project for testing, practice, or educational purposes 	To help you write your plan, DMPonline can show you guidance from a variety of organisations.
Project abstract	Select up to 6 organisations to see their guidance.
	Digital Curation Centre
The overall objective of MobiDataLab is to propose to the mobility stakeholders (transport organising authorities, operators, industry, innovators) a replicable methodology and	Find guidance from additional organisations below
sustainable tools that foster the development of a data sharing culture in Europe and beyond. MobiDataLab is based on a continuous co-development of knowledge and of technical solutions for data sharing with high involvement of all data producers and consumers in the transport and mobility landscape. This will be put in action through problem-solving oriented Labs, the collection and analysis of advice and recommendations of experts and supporting cities/regions/clusters/associations aided by the incremental	See the full list Save

Figure 3: MobiDataLab page on DMPOnline

3.4. Transport Cloud and Open Knowledge Base

The MobiDataLab Transport Cloud is a cloud-based prototype platform for sharing transport data, accessible to interested mobility actors.

This platform, technically designed according to federated cloud principles, shows how to facilitate access to mobility data in an open, interoperable and privacy-preserving way, using open tools. In particular, the MobiDataLab Transport Cloud aims to make mobility data FAIR by:

- allowing the discovery of data (either static or dynamic),
- providing access to mobility data,
- prototyping data processors to add value to the data and,
- demonstrating anonymisation and privacy-preserving tools.

To showcase that the project meets these four objectives, four demonstrators have been delivered as part of the prototype implementation (WP4), each one of them in two versions.

The Transport Cloud is primarily designed to demonstrate and offer solutions to reduce and, in some cases, remove current technical limitations identified as barriers to data sharing and reuse.





The Transport Cloud has used the Comprehensive Knowledge Archive Network (CKAN)¹⁷ and GeoNetwork¹⁸ as Data Catalogue demonstrators.

All information related can be found in D4.1 and D4.2 Transport Cloud Architecture Dossier, D4.3 and D4.4 Reference Data Catalogue.

The Open Knowledge Base (OKB) goal is to provide access to mobility data to the public and establish a common centre of reference. That way cities can, for instance, monitor their data as well as analyse what is being done elsewhere.

The OKB has been available to the public since the first trimester of 2023 and was initiated with the existing knowledge base of MobiDataLab. Anyone will then be able to contribute and enrich the base by adding articles or updating the ones already available.

3.5. Open data use

This part of the MobiDataLab DMP reports on open data used by the MobiDataLab project.

The respective datasets available are on the Transport Cloud CKAN (115 774) and GeoNetwork (199 099). A Reference Data Catalogue inventory text file provides details about the available organizations' data portals according to city/municipality/region and other criteria such as country, organization, description, themes, keywords, data formats, licenses, access policy, etc.

Such datasets encompass datasets that were collected from public authorities and institutions.

These datasets are released under an open license that allows us to use the data for research and non-commercial purposes.

The MobiDataLab's Transport Cloud Data Catalogue provides some documentation related to the usage of mobility data with open data software¹⁹ and open libraries²⁰.

The actual data can also, in general, be downloaded directly from the websites of respective organizations e.g., the "Open Data Roma Capitale" website. Although some of the data could also be stored on the MobiDataLab repositories for further processing and visualization, the respective CKAN resource (data) associated with the CKAN dataset (meta-data) links to the source of the dataset.

Thanks to the Transport Cloud Data Catalogue, companies or actors of public transportation can easily access the metadata of interest, therefore, gaining insights into their business.

²⁰ <u>https://github.com/MobiDataLab/CKAN-access-with-Jupyter-Notebook</u>





¹⁷ <u>https://ckan.mobidatalab.eu/</u>

¹⁸ <u>https://geonetwork.mobidatalab.eu/GeoNetwork-4.0.5-0/srv/eng/catalog.search#/home</u>

¹⁹ <u>https://github.com/MobiDataLab/QGIS-guide-CKAN-GeoNetwork-access</u>

3.6. Non-open data used

This part of the MobiDataLab DMP reports on non-open data used by the MobiDataLab project.

The respective datasets are going to be stored in a private repository and will be assigned the tag input-data (in addition to the respective WP tags) and associated with a group "non-open data used by the MobiDataLab".

3.7. Open data created

This part of the MobiDataLab DMP reports **open data** created by the MobiDataLab project.

The respective datasets have been put on Zenodo, linked to the OpenAire project page and are also accessible on the Open Knowledge Base for the documentation part, and on the Transport cloud for more Technical Data.

3.8. "Non-open" data created

This part of the MobiDataLab DMP reports on non-open data created by the MobiDataLab project.

The respective datasets are going to be stored in a private repository and will be assigned the tag input-data (in addition to the respective WP tags) and associated with a group of "non-open" data created by the MobiDataLab.





4. FAIR data principle

The acronym "FAIR" refers to the characteristics of data or metadata being Findable, Accessible, Interoperable and Reusable. The FAIR approach refers to a concise, domain-independent, highlevel and measurable set of guiding principles and practices to apply to a wide range of scientific data or metadata. However, they are not themselves a standard or a specification. Precisely, they precede implementation choices and do not necessarily suggest any specific implementation solution.

These principles ensure that the most important components for data lifecycle management are covered by data implementers, publishers and managers to evaluate whether their implementation choices are rendering their digital research artefacts FAIR. They form the basis of long-term care of valuable digital assets composed of the data produced by the research project while keeping the goal of being discovered and re-used by further research²¹.

The FAIR principles²² and practices are the result of the work in 2014 of a community of stakeholders representing academia, industry, funding agencies, and scholarly publishers, which were adopted the same year by the European Commission as the data guidelines for the Horizon 2020 (H2020) framework programme. These "emphasize on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals"²³.

In practice, the elements of the FAIR principles are related but independent and separable. Any combination of the principles can be applied incrementally. Thus, this modularity of the principles, as well as their distinction between data and metadata, facilitates their support in a wide range of special circumstances. The FAIR principles can also be applied to non-data assets which need to be identified, described, discovered, and reused in the same manner as data.

It is important to mention that there is a clear distinction between FAIR data and Open data, the former does not necessarily imply the latter.²⁴ While the openness of data is encouraged within the H2020 programme, there are necessary and legitimate reasons to restrict access to certain data. Nonetheless, the FAIR principles can still apply equally to restricted data or internal data of an organization, to make them more usable and valuable.

²⁴ https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_0.pdf





²¹ <u>https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm</u>

²² https://www.go-fair.org/fair-principles/

²³ Wilkinson, D. et al (2016), The FAIR Guiding Principles for scientific data management and stewardship [Online] Sci Data 3. Available from https://www.nature.com/articles/sdata201618 [15 March 2016]

Following the principle of "as open as possible, as closed as necessary"²⁵, research data should be open by default while setting a variable degree of openness. As illustrated by Figure 4, the more the data becomes both open and FAIR, the higher the benefits it provides.



Figure 4: The relationship between FAIR and Open data

In general terms, the research data produced within H2020 should be FAIR. The MobiDataLab project has built an EU-wide open data platform for searching and fusing multimodal mobility data, based on open-source projects/platforms, data channels for accessing the data both by humans and applications (Representational State Transfer (REST) Application Programming Interface (API)), and a toolbox for dataset enrichment and quality raising, containing so-called processors (Semantic, geographical, data format translations, privacy and anonymization mechanisms). As specified in the Exploitation plan deliverables D6.8 and D6.9 produced in T6.6, AKKODIS, as the Exploitation manager of MobiDataLab, oversees the Research Data Management, including the respect of the application of FAIR principles and the constant update of referencing platforms like Zenodo²⁶. The DMP reflects the decisions that affect the FAIRness of the data produced by MobiDataLab. In this third iteration, the specification and documentation of the actions taken are being updated.

In the following sub-sections, the guiding principles of each one of the four concepts of FAIR are cited. It also details how MobiDataLab has fulfilled the requirements of FAIR principles.

4.1. Making MobiDataLab (meta)data findable

- F1. (Meta)data are assigned a globally unique and persistent identifier.
- F2. Data are described with rich metadata (defined by R1 below).
- F3. Metadata clearly and explicitly include the identifier of the data it describes.
- F4. (Meta)data are registered or indexed in a searchable resource.

The MobiDataLab approach is based on three pillars: The Open Knowledge Base, the Transport Cloud and the Labs (Living and Virtual labs). The data produced or used in the project is made openly available by default, apart from the non-open data.

²⁶ https://zenodo.org/





²⁵ https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-datamanagement/data-management_en.htm

The deliverables uploaded in Zenodo will be assigned a Digital Object Identifier (DOI), making them citable, trackable, and thus Findable.

Additionally, naming conventions have been applied to all the deliverables, for which it is necessary to indicate in the following order: <Project Name>, <Deliverable, Task, or WP Code>, <Title> and <Version number>.

- The Deliverable, Task, or WP Code as defined in the Grant Agreement.
- The title of the document is named as in the Grant Agreement or abbreviated if it is too long (it should be written using upper cases).
- The version will have the format v'x'.'y', where 'x' and 'y' are progressive numbers.
- Whenever there is a new version of an article, deliverable, or demonstration, this is written.

In the case of the Open Knowledge Base:

- The OKB provides a web-based tool for searching by querying a large set of resources linked to challenges and emerging principles for data sharing.
- The OKB provides a URL to the different topic's pages.
- A table of contents is available in each document to provide the reader with a global idea of the content.

In the case of the Transport Cloud:

- MobiDataLab developed a prototype platform for searching, accessing and fusing multimodal mobility data in the cloud. This platform provides effective tools to facilitate access to and exchange of mobility data (by humans and machines) for mobility actors, data innovators and solution providers. The reference data catalogues allow the discovery of static, dynamic, realtime, and historical data for procedures, research, and/or data analysis; thus, making data findable.
 - The standards for metadata will be provided for each dataset if applicable as presented.
 - The datasets have metadata and a globally unique and persistent identifier to facilitate the findability.
 - The datasets are described by metadata with a plurality of relevant attributes.
 - Most of the datasets have a clear and accessible data usage license. In the case that this is not provided (after the harvesting process) directly on the metadata page, the license of the metadata source organization will be provided on the organization's page.
 - Datasets are associated with detailed provenance.
 - Except some of the datasets provided by the MobiDataLab reference group members²⁷ providing challenges²⁸ and partners of the consortium (who provided privileged data in the context of the Labs without license details, but who were aware that these datasets were going to be openly shared, unless precise otherwise by them), the datasets come from open data sources.
 - Most of the datasets meet domain-relevant community standards.
 - Open data formats: CSV, HTML, JPEG, JSON, RDF, PNG, PDF, tar, XML, ZIP.

²⁸ <u>https://labs.mobidatalab.eu/challenge-details/</u>





²⁷ <u>https://mobidatalab.eu/about/reference-group/</u>

- Mobility data standards: DATEX II, GTFS, GBFS, NeTEx, SIRI, TN-ITS.
- Geodata standards: ESRI REST, GEOJSON, GDB, GeoPackage, GeoDCAT-AP, KML, KMZ, MAP, OSM, PBF, SHP, TIFF, WFS, WMS, WMTS.
- Metadata exchange standards: CSW, INSPIRE, DCAT, DCAT-AP.
- Semantic interoperability standards: LOD, RDF, SPARQL.
- Other formats: ATOM, ERT, GML, JPG, MP4, MQTT, OAF, ODS, PPTX, RAR, RVC, STA, SQL, TSV, TXT, XLS, XLSX, XSD, XSL.

(The definition of most of these data formats is provided in the deliverables of WP2: State of the art on D2.4 Mobility data sharing standards and D2.5 Report on new Mobility data sharing standards)

- Most of the datasets are public. It is not necessary for discovering and accessing metadata, but
 registration is available to get notification of followed datasets when updated. Registration is for
 administrators and organizations publishing and personalizing features.
 - Only two datasets were marked private as the publisher requested not to make them visible after the living labs since the datasets will overcome major changes and it is preferable to avoid confusion. However, the two datasets will be available in the future as open access and harvested directly from the publisher's data portal, so they are only temporally restricted.
- Keywords will be provided in our repositories to optimize the possibilities for re-use.
- Metadata has been created to provide details about the datasets received by the reference group
 of local organizations proposing challenges during the Labs. The metadata was directly added
 to the data catalogues, and it is accessible through the graphical user interface and the API of
 the Data Catalogue.

In the case of the Labs:

- The Lab's challenges have a unique and persistent name and URL²⁹ to facilitate the findability.
- The datasets published in the virtual lab³⁰ redirect the user to metadata catalogues containing the globally unique and persistent identifier.
- A naming convention was defined for dataset file storage.

4.2. Making data openly accessible

A1. (Meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 The protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2. Metadata are accessible, even when the data are no longer available.

³⁰ <u>https://labs.mobidatalab.eu/dataset/</u>





²⁹ <u>https://labs.mobidatalab.eu/challenges/</u>

In compliance with the H2020 rules regarding Open Access (OA)³¹ to scientific literature, MobiDataLab will make any scientific publication accessible online for free under the scope of the project. The publications are deposited in a PDF format in the project's Zenodo repository (allowing discovery and rendering data accessible).

According to the publication type, the peer-reviewed scientific publications resulting from the project follow the most suitable approach among the two below:

- Self-archiving / 'Green' OA the author, or a representative, archives (deposits) the final peerreviewed manuscript, upon its publication approval or after, in an online repository. Access can be given after an embargo period has elapsed. If this is the case, the European Commission demands that open access is ensured within a maximum of six months.
- Open access publishing / 'Gold' OA an article is immediately published by the publisher; this can be as an online journal. In this model, open access must be granted at the latest on the date of publication. A copy of the publication should also be deposited in a repository.

The management of scientific information within H2020 under the open-access guidelines is illustrated as follows:



Figure 5: Open access to research data

In what regards the Open Knowledge Base:

- Most of the deliverables are made openly accessible on the website of MobiDataLab as well as on Zenodo.
- The data and their associated metadata are made available either in a public repository or in an institutional repository (MobiDataLab website).

³¹ https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf





In the case of the Transport Cloud:

- Guiding documentation and demonstrations have been published (to access and analyse the data) about the data enrichment processors^{32 33}, the metadata³⁴ and services³⁵ catalogues, the anonymizer³⁶ and the journey planner³⁷ have been made public on the <u>GitHub repository of MobiDataLab</u>. Registration is not required.
 - Open-source code about the proposed solutions is also available in this same repository.
 - This concerns also the Navitia source code of the Journey Planner V2³⁸; however, the source code of upcoming versions will not be open. In January 2023, there was a change in Hove's policies and the company decided to stop the open-source model for the Navitia Multicriteria calculator (Navitia V3). It is important to mention that Hove is proprietary for the data quality and processing brick for the offer repository, DataHub³⁹, and the disruption management traffic report software (both SaaS are closed). Now to request and have access to Hove services, in agreement with GDPR rules, there is a need to subscribe to a license. So, they have closed the open code for all products and restricted access to the open data bar on navitia.io and stopped the technical and functional support. This decision follows two main reasons: there was very little contribution to Navitia over the last ten years and an unviable business model for Hove.
- In terms of created data Navitia logs⁴⁰ were provided by Hove and origin-destination matrices⁴¹ were computed by URV from the data provided by Hove. The anonymization of the datasets was conducted by URV. Metadata about the anonymization process and Navitia measures were provided as documentation files. Navitia logs are proprietary, but Hove does not expose these logs as open data. Therefore, for these datasets, a Hove proprietary license applies.
- Navitia closed its open data bar, and no other solution exists currently. While the GTFS open data was freely available in an ODBL format, in its current state, it could not be marketed. Nevertheless, some of this data is accessible on the National Access Point of France, but European data is on an AWS repository managed by Hove's Data Management team, and no customer has open access for the moment.

In the case of the Labs:

- The data regarding the proposed Labs is accessible on the Virtual Lab website⁴². There is a particular deliverable explaining the functions of the Virtual Lab⁴³.
- The information on the challenges and datasets is available publicly. Registration (identification) is required in the Virtual Lab to identify the person and team submitting questions and/or solutions.

⁴³ https://mobidatalab.eu/wp-content/uploads/2023/01/MOBIDATALAB-D5.1-TheVirtualLab-v1.1DRAFT.pdf





Funded by the European Union

³² https://github.com/MobiDataLab/mdl-semantic-enrichment

³³ https://github.com/MobiDataLab/mdl-geo-enrichment

³⁴ https://github.com/MobiDataLab/reference-data-catalogue-demo

³⁵ https://github.com/MobiDataLab/mdl-catalog-api

³⁶ <u>https://github.com/MobiDataLab/mdl-anonymizer</u>

³⁷ https://github.com/MobiDataLab/mdl-datathon/blob/main/Navitia%20Playground%20ABC's_V3.pdf

³⁸ <u>https://doc.navitia.io/#getting-started</u>

³⁹ <u>https://navitia.com/solutions/data-factory/datahub/</u>

⁴⁰ https://ckan.mobidatalab.eu/dataset/trajectory-logs-datasets

⁴¹ <u>https://ckan.mobidatalab.eu/dataset/trajectory-logs-matrix</u>

⁴² https://labs.mobidatalab.eu/

MOBIDATALAB - H2020 G.A. No. 101006879

4.3. Making data interoperable

11. (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles.

I3. (Meta)data include qualified references to other (meta)data.

When considering the mobility sector in Europe it appears that there is a lack of interoperable data and insufficient guidelines for the usage of transport data exchange standards (e.g., Network timetable Exchange (NeTEx)) which are mainly used by European transport operators. Beyond public transport, few widely adopted standards exist for other modes of passenger transport as well. Facing these challenges, the Transport Cloud prototype platform has:

- Eased the aggregation and enriched data from the following standards, e.g., General Transit Feed Specification (GTFS)⁴⁴, Geographical features based on JSON (GeoJSON⁴⁵), Open Street Map (OSM⁴⁶).
- Triggered and provoked interoperable interfaces to connect data providers, service providers and transport clouds.
- Provide access to the data in multiple formats (bulk download, subsets of large datasets, etc.), thus, making data available through APIs (built on web standards such as CKAN API⁴⁷, GeoNetwork OpenAPI⁴⁸ and Explore API⁴⁹ completely documented and avoiding breaking changes), and enabling to deal with real-time data, providing data up to date / archived data, and making them Interoperable. The Catalogue Service for the Web (CSW) endpoint of GeoNetwork exposes the metadata records in XML format using the OGC CSW protocols⁵⁰, while the CKAN catalogue of MobiDataLab enables a function that allows CKAN to expose and consume metadata from other catalogues using Resource Description Framework (RDF)⁵¹.
- Employed the ISO 639-1 and IETF BCP 47 standard vocabularies for datasets added manually into the main Data Catalogue, to add information such as the country code or the alpha-2 code, including details about the content and/or the metadata language of the datasets.
- Used ontologies during the development of the semantic enrichment processor (for more information consult D2.5).
- Allocated an open license to most of the datasets in the data catalogues, which allows extensive re-use.
- Guaranteed the data produced and used during the project is useable by third parties, after the end of the project. As explained earlier, only restricted datasets will not be able to be re-used.
- The data is intended to remain usable for at least 3 years and the website will be available for the next 5 years.

⁵¹ <u>https://github.com/MobiDataLab/ckanext-dcat#rdf-dcat-endpoints</u>





⁴⁴ https://gtfs.org/

⁴⁵ https://geojson.org/

⁴⁶ <u>https://www.openstreetmap.org/</u>

⁴⁷ https://docs.ckan.org/en/2.9/api/

⁴⁸ https://geonetwork.mobidatalab.eu/GeoNetwork-4.0.5-0/doc/api/index.html

⁴⁹ <u>https://data.opendatasoft.com/api/explore/v2.1/console</u>

⁵⁰ https://geonetwork-opensource.org/manuals/4.0.x/en/api/csw.html

On the side of the Virtual Lab:

- The Virtual lab has an Open API. Interoperability between the Virtual Lab and the Transport Cloud is possible through an API and the Virtual Lab can upload and retrieve data or metadata directly from the Transport Cloud.
- A naming convention was defined for dataset files' storage.

It is also important to note that Zenodo ensures Interoperability by allowing GitHub integration to preserve metadata.

4.4. Increase data re-use

- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes.
 - R1.1. (Meta)data are released with a clear and accessible data usage license.
 - R1.2. (Meta)data are associated with detailed provenance.
 - R1.3. (Meta)data meet domain-relevant community standards.

To render data reusable, WP4 of MobiDataLab provided prototype data processors to add value to the data. To be precise, Task 4.4 started with the identification of datasets with missing values, which could greatly be enhanced by combining the data with other datasets and gathering additional results. Semantic and geographical data enrichment techniques have been provided as open tools. More specifically, this task aimed to:

- enrich data semantically (combining with the Linked Open Data (LOD) cloud, RDF⁵²/ SPARQL Protocol and RDF Query Language (SPARQL)⁵³),
- enrich data geographically (tools for geocoding/projection, cross-referencing with spatial datasets, geodata APIs, OSM data, etc.),
- enrich data through complementary presentations (maps, tables, RDFs, etc.),
- enrich data through machine learning and,
- verify that the enrichment does not distort the statistical power of the data.

The transport cloud prototype produced in WP4 was primarily designed to demonstrate and offer solutions to reduce and, in some cases, remove current technical limitations identified as barriers to data reuse. Since data is dynamic and requires constant maintenance, in the future the verification of the metadata provided in the data catalogues will be effectuated to ensure that the datasets are up-to-date and that data licenses are clear.

OpenAIRE, through repository harvesting and mining techniques, infers links between publications, research funding and research data, which enables data reuse. On the other hand, Zenodo versioning gives the possibility to update datasets easily and store all changes in metadata over the record's lifetime, which also enables reuse.

52 https://www.w3.org/RDF/

⁵³ https://www.w3.org/2001/sw/wiki/SPARQL





The results of the MobiDataLab project, i.e., data, knowledge or information produced and generated, are attached to the Intellectual Property Rights (IPR⁵⁴). In the absence of specification, national IPR applies. However, sometimes it is difficult to interpret certain national laws and the reuse of these results can be increased with a license. Therefore, all data producers should license their data to allow the widest reuse possible. The open research data will be licensed under the Creative Commons BY⁵⁵ 4.0 DEED Attribution 4.0 International license⁵⁶. To learn more about data licensing, please refer to deliverable 1.5 section 6.1.2. By default, the data is made available for reuse. If any constraints exist, an embargo period will be mentioned to keep the data from being reused for a limited period.

⁵⁶ https://creativecommons.org/licenses/by/4.0/





⁵⁴ <u>https://commission.europa.eu/business-economy-euro/doing-business-eu/intellectual-property-rights_en</u>

⁵⁵ BY: credit must be given to the creator (https://creativecommons.org/share-your-work/cclicenses/)

5. GDPR

Within the European Union (EU), during the early stages of the Internet, the 1995 Data Protection Directive was adopted. Over the last 25 years, tremendous changes in technology have brought the need to revise data protection regulations. In 2016, the EU adopted the General Data Protection Regulation. Now GDPR is recognized as a law across the EU.

GDPR has been enforced in 2018 and provides now broader data privacy for individuals and new obligations for any data-processing organization (data collector, publisher, implementor) - regardless of geographical location - that collects, uses, or processes EU citizens' personal information. Personal data security is thus improved via the enforcement of the four following aspects:

- Right to be forgotten: Article 17⁵⁷ mentions the obligation for the organization, and any business
 partners with whom they have shared data, to delete any personal data from their systems upon
 request.
- Data protection by design and default: Article 25⁵⁸ stipulates that organizations must set internal policies and measures to protect data by design and default, and all applications, services, and products must adhere to these policies.
- Secure data processing: Article 32⁵⁹ requires that organizations can prove an implementation of measures to ensure appropriate levels of security.
- Timely breach notifications: Article 34⁶⁰ imposes hefty fines on organizations if breaches of unencrypted data are not reported to authorities and affected individuals within 72 hours.

GDPR has been identified as one of the targets of MobiDataLab. The current context of the project shows a variety of factors bringing different barriers and obstacles for it to reach higher technology readiness levels. Amongst the legal factors, GDPR plays a data protection role and its "purpose limitation" principle (requirement that data is collected for one specified, explicit, and legitimate purpose, and be not processed for a new, incompatible purpose) can hinder our progress. For this reason, MobiDataLab has anticipated these limitations through the legal studies in WP2 and included an assessment of factors influencing data governance, including applicable legal regimes such as the GDPR.

5.1. Personal Data Collection, Processing, and Re-use

A limited number of partners of MobiDataLab have collected and processed personal data. In the context of the Labs, tracking of the participants and observations related to the solutions submissions has required processing. Therefore, MobiDataLab is subject to the provisions of the GDPR which enshrines the following key principles (without considering exemptions):

⁶⁰ <u>https://gdpr-info.eu/art-34-gdpr/</u>





⁵⁷ <u>https://gdpr-info.eu/art-17-gdpr/</u>

⁵⁸ https://gdpr-info.eu/art-25-gdpr/

⁵⁹ <u>https://gdpr-info.eu/art-32-gdpr/</u>

- Data must be processed fairly, lawfully and only for the purpose for which it was collected and further processed.
- Data cannot be disclosed without authorization unless there is an overriding act of law or legitimate grounds to do so.
- Subject to certain exemptions, individuals have a right to access the information relating to them and to ask for correction of inaccurate data.
- Information cannot be transferred beyond the European Economic Area boundaries without consent or adoption of other adequate protection measures.
- Organizations are usually required to register or notify the processing of personal data unless the data processing is simplistic, or a data protection officer has been appointed.
- Organizations must have adequate security measures in place.
- The MobiDataLab consortium exercises a principle of the minimum resort to notification exemptions afforded under national laws.

GDPR and Consent Information sheets were provided to the participants for each of the Labs of MobiDataLab to which they registered.

5.2. Compliance with GDPR Recital 78⁶¹

The data must be related to an identified or identifiable living individual. This individual does not need to be directly identifiable but may be identified by a reference number or some other tag which, in each small group or through analysis of patterns in sufficient volumes of data, might allow an individual to be singled out from a group. Based on the kinds of data sources to be included in this research, direct personal identifiers (e.g., specific names or faces) may exist in a variety of locations within the dataset.

MobiDataLab's default anonymization process(es) are 'one-way' with original source data disposed of in such a way that re-identification of data or decoding of anonymization tokens by reference to any 'real-world' datasets is rendered difficult to the greatest extent possible. MobiDataLab follows the guidance outlined in Article 29 Working Group 05/2014 Opinion on Anonymization Techniques⁶² and specifically its recommendations on pseudonymization, noise addition, substitution, aggregation, k-anonymity, l-diversity, differential privacy and hashing/tokenization. MobiDataLab also includes downstream contractual obligations as a legal measure to respect privacy in the use of the project's results.

5.3. Data Protection Agency Notification

The data to be processed in MobiDataLab is unlikely to constitute Personal Data within the meaning of the EU Directive and relevant national legislation.

⁶² https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf





⁶¹ <u>https://gdpr-info.eu/recitals/no-78/</u>

MobiDataLab is also likely to be exempted from national notification processes because the data collection is for scientific research and because of additional institutional data protection measures, access restrictions and more conditions put in place. The consortium is aware that anonymization approaches must be applied to video/still images within data to avoid the risk that a token identifier might become associated with sufficient unique data points to uniquely identify a living individual. The consortium undertakes to notify data protection authorities⁶³ in jurisdictions where research activities are carried out and specific relevant actions within it to obtain, if necessary, authorization for such activities. The exact requirements and due diligence need to be scoped and defined within the relevant jurisdictions.

The relevant national approvals are sought and acquired as and when necessary, as the notification requirements vary from one country to another and therefore no single timeline can be provided for the completion of all notification procedures⁶⁴. Renewal of notifications, when necessary, is carried out in line with the requirements of different national legislations. Processes for notification vary from one jurisdiction to another. The following project principals have been assigned the responsibility of acting as interlocutors with their national data protection agency. Further information on notification procedures and the relevant agencies in Europe can be found in Article 29 Working Group document "Vade mecum on Notification Requirements".

Renewal of notifications, when necessary, is carried out in line with the requirements of different national legislations. The renewal requirements are also leveraged to react to change in the project and the adoption of video archives into research, development and test tasks.

5.4. Compliance with Article 49⁶⁵ of the GDPR

While our position, as scientific researchers, permits us derogation from the prohibition on processing (sensitive categories of) personal data, the consortium is nevertheless aware that it remains incumbent upon us to provide specific and suitable safeguards to protect the fundamental rights and privacy of data subjects. Some of these precautions are already detailed above. MobiDataLab further undertakes to ensure that any personal data collected is treated under Article 49. Personal data collected will be processed fairly and lawfully. Personal data collected is used only for research purposes as specified in our original proposal. The data is adequate, relevant and not excessive concerning the purposes for which it had been collected.

The consortium endeavours not to collect, but to expunge, all data that is not directly project-related.

⁶⁴ https://ec.europa.eu/justice/article-29/documentation/other-document/files/2006/2006-07-03-vademecum.doc

⁶⁵ https://gdpr-info.eu/art-49-gdpr/





⁶³ <u>https://commission.europa.eu/law/law-topic/data-protection/reform/what-are-data-protection-authorities-dpas_en</u>

5.5. Safe Harbour and EU–US Privacy Shield Considerations

Considering the European Court of Justice decisions on the International Safe Harbour agreement and the EU–US Privacy Shield (October 6th, 2015, and July 16th, 2020)⁶⁶, the MobiDataLab project stores all data derived from personal data (after anonymization or dissociation) in EU member states and complies with Article 29 Working Group⁶⁷ communique on the transfer of data outside of the union and forthcoming member state decisions on the Trans-Atlantic Data Privacy Framework⁶⁸.

5.6. Activities Dedicated to Ethical Considerations

The guiding principles at the heart of the MobiDataLab approach are the highest ethical standards, the protection of privacy and the validity of data and its accurate representation. In adhering to these principles and remaining cognizant of concerns that arise in the work plan, MobiDataLab has tested and evaluated research results in WP3 under the principle of informed consent. Consent information was included in questionnaires and registration forms dealing with personal data for data sharing and long-term preservation. The consent forms provided during the Labs are available on the website of the project:

- Datathon
- Hackathon
- <u>Codagon</u>

In addition to the measures mentioned above, MobiDataLab takes the following steps toward addressing these:

- Follow the ethics and data protection guidelines⁶⁹ made by the Directorate-General for Research and Innovation experts.
- Ensure to be constituted of available partners with Ethics, Privacy and Legal expertise to all MobiDataLab project staff members at the outset of the project and throughout.
- Assurance that privacy by design⁷⁰, ethics, legal and societal impact requirements are included as research and development mandates integrated into the MobiDataLab project plan in compliance with GDPR Article 25.

https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf





⁶⁶ <u>https://en.wikipedia.org/wiki/EU%E2%80%93US_Data_Privacy_Framework#cite_note-1</u>

⁶⁷ https://edpb.europa.eu/system/files/2023-09/wp260rev01 en.pdf

⁶⁸ https://ec.europa.eu/commission/presscorner/detail/en/FS 22 2100

⁶⁹ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-dataprotection_he_en.pdf

5.7. Minimum Resort to Exceptions and Derogations

The GPDR allows for exceptions and derogations for personal data used for research. For example, general exemptions for processing certain categories of sensitive personal data (e.g., Article 6⁷¹ and Recital 50⁷²). Exceptions for a right to opposition for processing or storage of data (Article 89), and processing of data without consent (Article 6.1.f, Recitals 47⁷³ and 157⁷⁴), may be applicable. MobiDataLab commits to a minimum resort to exceptions and derogations in the processing of personal data within the project for research purposes.

5.8. Activities Dedicated to the Protection and Securing of Personal Data

As project coordinator, AKKODIS has ensured that the consortium guarantees the treatment of personal data generated during the project. This is done via a set of development directives and methodologies, to ensure the project applies adequate database encryption and secure systems techniques.

Furthermore, the directives and mandates also integrate the technical requirements of European, national and regional data protection legislation. Partners are required to have adequate security measures in place, both technical (firewalls, access controls, access audits, etc.) and operational (training, incidence reporting, etc.). The following range of issues is considered in establishing such directives and mandates:

- categories of sensitive data,
- security measures for sensitive data,
- policies for fair acquisition and processing of data,
- data retention policies,
- legal basis for the information processing,
- policies for processing compatible with the purpose,
- policies for data controllers and data processors,
- description of the technical characteristics of the processed data,
- technical features and topology of the information systems where data is stored and processed.

Further consideration shall be given to the following relevant regulations, decisions and guidelines. The following EU regulations are recognized to be relevant to the project:

• The Charter of Fundamental Rights⁷⁵ of the EU (particularly articles 3, 7, 8 and 25).

⁷⁵ <u>https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT</u>





⁷¹ https://gdpr-info.eu/art-6-gdpr/

⁷² https://gdpr-info.eu/recitals/no-50/

⁷³ https://gdpr-info.eu/recitals/no-47/

⁷⁴ https://gdpr-info.eu/recitals/no-157/

- Convention for the Protection of Individuals about automatic processing of personal data (Convention Nr. 108)⁷⁶.
- WMA Declaration of Helsinki⁷⁷ (especially articles 13, 20, 21, 22, 23 and 24, 25).
- Convention of the Council of Europe on Human Rights and Biomedicine (signed in Oviedo on 4 April 1997)⁷⁸.
- CIOMS -- International Ethical Guidelines for Biomedical Research Involving Human Subjects⁷⁹.
- Code of Nuremberg (Article 10)⁸⁰.

5.9. Shared Information and Personal Data

The partners agree that any background, results, confidential information and/or all data and/or information that is provided, disclosed or otherwise made available between the partners during the implementation of the action and/or for any exploitation activities ("shared information"), shall not include personal data as defined by under Article 4.1⁸¹ of the GDPR. Accordingly, each partner agrees that it will take all necessary steps to ensure that all Personal Data is removed from the Shared Information, made illegible, or otherwise made inaccessible (i.e., deidentify) to the other partners before providing the shared information to such other partners.

⁸¹ https://gdpr-info.eu/art-4-gdpr/





⁷⁶ <u>https://www.coe.int/en/web/data-protection/convention108-and-protocol</u>

⁷⁷ https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/

⁷⁸ https://www.coe.int/en/web/bioethics/oviedo-convention

⁷⁹ https://cioms.ch/wp-

content/uploads/2016/08/International Ethical Guidelines for Biomedical Research Involving Human Subjects.pdf ⁸⁰ https://en.wikipedia.org/wiki/Nuremberg Code

6. Allocation of resources

In the project, AKKODIS plays the role of data manager and liaises with the Executive Committee of the project about data management issues. The Data Manager leads Data Management Plan tasks and participates in the project coordination monitoring the evaluation of data collection, storage and handling, as well as their publication as part of the ORD Pilot.

All research data collected as part of this project is owned by the data producer or partners involved in the Virtual and Living Labs. The partners in MobiDataLab take responsibility for the collection, management, and sharing of the research data. Quality assessment falls under the responsibility of the data manager of each Virtual and Living Lab.

The costs to make the data FAIR in MobiDataLab shall be handled by each partner who has to generate its data according to the requirements expressed in the Data Management Plan. The different tasks of WP4, especially, contribute to fulfilling the requirements of the FAIR principles.

First, Task 4.2 "Reference Data Catalogue" provides a state-of-the-art open data catalogue referencing transport datasets and corresponding metadata in the territorial context and specific domains of the "Reference Group" of MobiDataLab stakeholders. The resulting catalogue uses explicit dataset descriptive information (title, description, keywords, publication date, data source, spatial coverage, media type, etc.) allowing both human understanding and automatic discovery by software agents, thus making data findable. This task is led by AKKODIS with a resource effort of 11 person-month (PM) distributed from month 6 to month 34 of the project.

Second, Task 4.3 "Data Access Services and Data Channels" performed the practical integration of available data access services and data channels. Availability, documentation and usability have been key assets of this part. Aligned with the work done in the other Transport Cloud work packages, within this task, available datasets and services from project partners and their platforms are connected and have become accessible. The activities are connected to parallel tasks in this context and have generated positive impact. HERE is the Task Leader and has made an effort of 4 PM for the period from month 6 to month 34 of the project.

Third, MobiDataLab data has been made interoperable by two tasks. On one hand, via Task 4.3 "Data Access Services and Data Channels", the Transport Cloud prototype has provided access to the data in multiple formats (bulk download, subsets of large datasets, etc.), making data available through APIs (built on web standards, completely documented and avoiding breaking changes), able to deal with real-time data, providing data up to date / archived data, and also making them interoperable. On the other hand, Task 4.4 Data Processors have contributed to the development of open tools allowing the enrichment of data. For example, tools for spatial enrichment include geocoders to convert a human-readable address into spatial coordinates, geo converters to simplify, convert or normalize geographical data (e.g., projections) and spatial functions to run computations (e.g., distance computation) on them. Such data format translations make data interoperable. AKKODIS is the Leader of Task 4.4 which manages 7 PM on this task from month 6 to month 34 of the project.





Fourth, Task 4.4 Data Processors starts with the identification of datasets with missing values, which can greatly be enhanced by combining the data with other datasets and gathering additional results. Different data enrichment techniques, including semantic enrichment and geographic transformations, are provided as open tools, generating new data and encouraging *Reuse* by adding value. Moreover, Task 4.5 Anonymization and Data Privacy includes data processing modules that apply data protection and anonymization techniques to the mobility data in the Transport Cloud. New protection techniques have been researched and developed where the current state-of-the-art techniques were not able to fulfil the project requirements, thus promoting high-value data reuse. Task 4.5 Leader is URV, who has at its disposal 18 PM throughout the task period running from month 6 to month 34 of the project.





7. Data security

7.1.1. Storage

In the European framework, the European Open Science Cloud (EOSC) is an initiative providing a world-class data infrastructure to store and manage data. EOSC aims to solve the problem of fragmented access and non-interoperable research data centres across Europe. In terms of data storage, EOSC provides several reliable, secure and scalable cloud storage solutions for scientific data, apps and workloads. Amongst them, some are open access or even fully open access.

EOSC provides also other services such as Open Science publishing infrastructure which is connected to the EOSC. OpenAIRE is an example which contributes actively to EOSC.⁸²

As mentioned in the previous section concerning Open access, MobiDataLab partners use such an open-access repository, as OpenAIRE, to grant access to the publications and bibliographic metadata in a standard format.

7.1.2. Authentication and authorisation

During their storage, data should be protected against any type of modification by the implementation of some security principles. The security principles are listed below:

- **Authentication**: All the users wanting to get access to the MobiDataLab project repositories should be authenticated. Furthermore, proper means are used to authenticate the repositories. An authentication system is used to handle the authentication of the users during the project.
- **Authorization**: Access to MobiDataLab project repositories is only available to authenticated and authorized users. These categories and the rights of those users are defined and enforced. The appropriate access control policies and mechanisms (including physical access control) shall be identified for the relevant task and project-wide to provide the authorization.
- **Accounting**: In MobiDataLab, any access and modification to a resource by any user is securely logged to prevent users from denying that data files were accessed, altered or deleted. Other accounting mechanisms shall be implemented.
- **Confidentiality**: The data stored in MobiDataLab repositories should be encrypted during transmission and storage.
- **Communication Security**: Access to MobiDataLab repositories should be done through encrypted communication channels such as Hypertext Transfer Protocol Secure (HTTPS) and Internet Protocol Security (IPsec).
- **Data Integrity**: The data collected during MobiDataLab should be protected from malicious and accidental modifications by any users during their transmission or their storage. Cryptographic mechanisms such as hash functions and digital signatures shall be used.

⁸² https://www.openaire.eu/openaire-and-eosc





 Availability: This security principle assures that the MobiDataLab repositories should be available for MobiDataLab users during the defined interval of service. Also, regular backups of the data should be made. Therefore, mechanisms to cope with the charge and Denial-of-service attacks should be implemented.

7.1.3. Data protection

Securing stored digital data involves preventing unauthorized people from accessing it as well as preventing accidental or intentional destruction, infection or corruption of information. While data encryption is a popular mechanism, it is just one of many techniques and technologies that can be used to implement a tiered data security strategy. Steps to secure data involve understanding applicable threats, aligning appropriate layers of defence and continual monitoring of activity logs acting as needed. This means that a multi-tier approach needs to be adopted by all the partners.

The proper method of storage and the appropriate community along with levels of access for privileged users are important considerations for comprehensive protection. Improperly stored information along with overly permissive accounts is a centralized theme in many high-profile breaches. Partners within MobiDataLab follow a specific set of guidelines to comply with the project's main requirement for the storage of digital data.

These are:

- Data availability must be guaranteed.
- Confidential data must be stored using access protection.
- Strictly confidential information must only be stored in an encrypted mode.
- Confidential data must not be stored in online services that are not approved by the MobiDataLab consortium.
- Any exception from this measure must explicitly be approved.
- Modifications to data with high integrity requirements must be documented and approved by the partners.





8. Conclusion

This deliverable covered the third and last version of the Data Management Plan of the MobiDataLab project.

The first version of this document provided an overview of the data processed in MobiDataLab and described the data categories of the project, blinding information on their management and the implementation of the FAIR principles in the project. As mentioned in the Grant Agreement, MobiDataLab is fully committed to the principles of the H2020 ORD Pilot, consequently, guidelines associated with open access were described to ensure that the project's outputs become openly available to the research community.

The second version of the Data Management Plan provided details regarding the description of the data handled by the project, the ORD Pilot data access, licensing, and GDPR. Corresponding standards and methodologies were provided, as well as how to curate and preserve the resulting data of the project.

The current version presented details and links towards the data and metadata handled and produced throughout the project. Great consideration is given to mobility datasets shared during the Living and Virtual Labs (WP5). Since different categories of data require different treatments, the definition and appropriate methodologies to handle different data/metadata were given. In the same way, the data from our reference group of stakeholders was detailed and their specificities duly described.

This deliverable shows how MobiDataLab addressed the two pillars⁸³ of ORD Pilot by establishing and maintaining a DMP (as a deliverable), by providing open access to research data as much as possible. This data includes public deliverables, non-protected results or results coming after a patent registration, which have been produced throughout the whole project's duration and shared in a specific part of the project's website (tailored to different levels of internal and external stakeholders), but also published on the general-purpose open repository, Zenodo. Furthermore, this deliverable provides examples all through on how MobiDataLab complies with the FAIR principles.

⁸³ <u>https://www.openaire.eu/what-is-the-open-research-data-pilot</u>





9. References

Data protection working party & European Commission. (2018a). *Guidelines on Personal data breach notification under Regulation 2016/679.* https://edpb.europa.eu/system/files/2023-09/wp260rev01_en.pdf

Data protection working party & European Commission. (2018b). Guidelines on Personal databreachnotificationunderRegulation2016/679.https://www.cnil.fr/sites/cnil/files/atoms/files/wp250rev01_en-data-breach.pdf

Directorate-General for Research and Innovation & European Commission. (2021, May 7). *Ethics and data protection.* https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

European Commission. (n.d.). Data management—H2020 Online Manual. European Commission. Retrieved 4 January 2024, from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

European Commission. (2022, March). Trans-Atlantic Data Privacy Framework. EuropeanCommission-EuropeanCommission.https://ec.europa.eu/commission/presscorner/detail/en/FS_22_2100Commission

European Data Protection Board. (2023). *Guidelines 9/2022 on personal data breach notification under GDPR. Version 2.0. Adopted 28 March 2023.* https://edpb.europa.eu/system/files/2023-04/edpb_guidelines_202209_personal_data_breach_notification_v2.0_en.pdf

European Parliament. (n.d.). *Data Protection Statement—On-premise Web Audio and Video Conferencing Service*. Retrieved 15 January 2024, from https://www.europarl.europa.eu/data-protect/reportPdf/printPrivacyStatement.do;jsessionid=20A7C5AA16FAFD95184DFA17049923BA ?prefix=V3&nr=180

European Research Council. (2017). Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

Experts Working Group on data protection and privacy. (2009). *Data protection and privacy ethical guidelines*. https://ec.europa.eu/research/participants/data/ref/fp7/89827/privacy_en.pdf

FAIR Principles. (n.d.). GO FAIR. Retrieved 4 January 2024, from https://www.go-fair.org/fair-principles/

How to GO FAIR. (n.d.). GO FAIR. Retrieved 4 January 2024, from https://www.go-fair.org/how-to-go-fair/

Koulocheri, E. (2017, November 22). *What is the EC Open Research Data Pilot?* OpenAIRE. https://www.openaire.eu/what-is-the-open-research-data-pilot





Phil Archer, Nikos Loutas, Stijn Goedertier, & Saky Kourtidis. (2012). *Study on persistent URIs.* https://philarcher.org/diary/2013/uripersistence/





MobiDataLab consortium

The consortium of MobiDataLab consists of 10 partners with multidisciplinary and complementary competencies. This includes leading universities, networks and industry sector specialists.





in https://www.linkedin.com/company/mobidatalab

For further information please visit www.mobidatalab.eu



MobiDataLab is co-funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).

The content of this document reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein. The MobiDataLab consortium members shall have no liability for damages of any kind that may result from the use of these materials.





