

# MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

## **| D1.4 Data Management Plan (1st issue)**

23/01/2023

Author(s): Chris WONG, Thierry CHEVALLIER, Xavier VALENTE



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

## Summary sheet

<b>Deliverable Number</b>	D1.4
<b>Deliverable Name</b>	Data Management Plan (1st issue)
<b>Full Project Title</b>	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
<b>Responsible Author(s)</b>	Chris WONG (AKKA)
<b>Contributing Partner(s)</b>	AETHON, AKKA, CNR, F6S, HERE, ICOOR, HOVE, KUL, POLIS, URV
<b>Peer Review</b>	HERE, URV
<b>Contractual Delivery Date</b>	31-07-2021
<b>Actual Delivery Date</b>	27-07-2021
<b>Status</b>	Final
<b>Dissemination level</b>	Public
<b>Version</b>	V1.0
<b>No. of Pages</b>	31
<b>WP/Task related to the deliverable</b>	WP1/T1.4
<b>WP/Task responsible</b>	AKKA/AKKA
<b>Document ID</b>	MobiDataLab-D1.4-DataManagementPlanV1-v1.0
<b>Abstract</b>	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

Organisation	Country	Abbreviation
AKKA I&S	France	AKKA
CONSORZIO INTERUNIVERSITARIO PER L'OTTIMIZZAZIONE E LA RICERCA OPERATIVA	Italy	ICOOR
AETHON SYMVOULI MICHANIKI MONOPROSOPI IKE	Greece	AETHON
CONSIGLIO NAZIONALE DELLE RICERCHE	Italy	CNR
HOVE	France	HOVE
HERE GLOBAL B.V.	Netherlands	HERE
KATHOLIEKE UNIVERSITEIT LEUVEN	Belgium	KUL
UNIVERSITAT ROVIRA I VIRGILI	Spain	URV
POLIS - PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES	Belgium	POLIS
F6S NETWORK IRELAND LIMITED	Ireland	F6S

## Document history

Version	Date	Organisation	Main area of changes	Comments
0.1	31/03/2021	AKKA	Table of content	Initial version
0.2	29/06/2021	AKKA	all	Draft version
0.3	09/07/2021	HERE, URV	all	Peer Review
0.4	20/07/2021	AKKA	all	Rework
1.0	27/07/2021	AKKA		Quality check and submission

## Executive Summary

Defining how to handle research data during and after the project is the topic of the Task 1.4 of the project (Data Management) and of the following deliverable (Data Management Plan v1). More specifically, this document identifies the data to be collected, processed or generated; what methodologies and standards will be followed (namely the FAIR principles to make data Findable, Accessible, Interoperable and Reusable on the one hand, and the General Data Protection Regulation on the other). In addition, the following deliverable describes how the data will be shared and/or made open, and how it will be curated and preserved.

## Table of contents

1. INTRODUCTION.....	8
1.1. PROJECT OVERVIEW.....	8
1.2. PURPOSE OF THE DELIVERABLE.....	8
1.3. INTENDED AUDIENCE.....	8
1.4. STRUCTURE OF THE DELIVERABLE AND ITS RELATION WITH OTHER WORK PACKAGES/DELIVERABLES.....	9
2. DATA HANDLED BY THE PROJECT.....	9
2.1. ADMINISTRATIVE DATA.....	9
2.2. OPEN RESEARCH DATA.....	10
2.3. EVALUATION DATA.....	11
2.4. TECHNICAL DATA.....	11
3. METHODOLOGIES AND STANDARDS.....	14
3.1. FAIR PRINCIPLES.....	14
3.1.1. Making data Findable.....	15
3.1.2. Making data Accessible.....	16
3.1.3. Making data Interoperable.....	16
3.1.4. Making data Reusable.....	17
3.2. GDPR.....	18
3.2.1. Personal Data Collection, Processing, and Re-use.....	18
3.2.2. Appointment of Data Protection Officer.....	19
3.2.3. Compliance with GDPR Recital 78.....	19
3.2.4. Data Protection Agency Notification.....	20
3.2.5. Compliance with Article 49 of the GDPR.....	20
3.2.6. Safe Harbour and Privacy Shield Considerations.....	21
3.2.7. Activities Dedicated to Ethical Considerations.....	21
3.2.8. Minimum Resort to Exceptions and Derogations.....	21
3.2.9. Activities Dedicated to Protection and Securing of Personal Data.....	22
3.2.10. Shared Information and Personal Data.....	23
4. ALLOCATION OF RESOURCES.....	23
5. ACCESS AND CURATION.....	24
5.1. DATA ACCESS.....	24
5.1.1. Open Access and ORDP.....	24
5.1.2. Data Licensing.....	26
5.1.3. Storage.....	27
5.1.4. Authentication and authorisation.....	28
5.2. DATA QUALIFICATION AND CURATION.....	28
5.2.1. Data protection.....	28
5.2.2. Data qualification.....	29
6. CONCLUSION.....	29

## List of figures

Figure 1 The relationship between FAIR and Open data.....	15
Figure 2 Open access to research data .....	26

## List of tables

Table 1 Dataset description template .....	12
--	----

## Abbreviations and acronyms

Abbreviation	Meaning
API	Application Programming Interface
CC	Creative Commons
D	Deliverable
DMP	Data Management Plan
EC	European Commission
EOSC	European Open Science Cloud
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
H2020	Horizon 2020
KPI	Key Performance Indicator
M	Month

MaaS	Mobility as a Service
OA	Open Access
OKB	Open Knowledge Base
ORDP	Open Research Data Pilot
OSI	Open Source Initiative
PM	Person-Month
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 as well as 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 will be 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 and how they are processed and preserved. It describes the approach to making the MobiDataLab data Findable, Accessible, Interoperable and Reusable (FAIR) 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 (ORDP) requirements. The document will also address ethical and confidentiality issues and some data security principles. This deliverable is a living document and will be updated as the project evolves:

- The initial version D1.4 (M6, i.e. July 2021) describes 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) will provide an update of the Data Management Plan including the description of the datasets (in particular the mobility data used in the project) and the potential evolution of the procedures defined at the beginning of the project;
- The final version D1.6 (M36, i.e. January 2024) will include the final description of the datasets and procedures.

## 1.3. Intended audience

The dissemination level of this D1.4 deliverable is 'public' (PU). AKKA as WP1/Task 1.4 leader is responsible for it, and its main contributor, with the assistance of all partners of the consortium. Appointed peer reviewers are HERE and URV.



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

This deliverable serves as an entry point to understanding the project-wide approach to data management in MobiDataLab. It provides an overview of data management at project level with a stronger focus on research data management, as required by the H2020 ORDP programme, even though later versions will also cover categories of mobility data that will be used in the course of the project.

As an input, it provides links to and between 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 will be 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), which aims to make the mobility data FAIR (Findable, Accessible, Interoperable and Reusable) 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. Data Handled by the Project

MobiDataLab will handle different types of data which can be organised in four categories: administrative data, open research data, evaluation data and technical data. This chapter will describe these data categories while providing information on their management and sharing.

### 2.1. Administrative data

This category refers, on one hand, to the data produced by the project management activities within WP1 such as meeting minutes, recordings, internal reports, for historical purposes and follow-up. In terms of deliverables, they include D1.1 “Collaborative portal”, D1.2 and D1.3 “Project, Innovation and Quality management plan”. The data are collected by the management team including the project manager, the WP leaders and task leaders. They 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 in order to access to all available functionalities linked to the project. These administrative data are confidential and are available only for the members of the consortium.

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

- Data management: D1.4, D1.5 and D1.6 “Data Management Plan”;

- Data exploitation: D6.8 and D6.9 “Exploitation Plan”;
- Project activities communication: D6.1 “Dissemination report”, D6.2, D6.3 and D6.4 “Reporting on MobiDataLab events”;
- Activities concerning project stakeholders: Specifically, there are deliverables D1.7 “Report on Expert Committee activities”, D6.5 “Stakeholder group activities report”, D6.6 and D6.7 “Project cooperation activities report”.

These data are open to the public and are accessible via the project website ([www.mobidatalab.eu](http://www.mobidatalab.eu))

## 2.2. Open Research data

The open research data category refers to the data resulting from the research work. This category includes mostly the deliverables from WP2 and WP3.

WP2 will consolidate an Open Knowledge Base derived from the most important projects and initiatives implemented to date in the domain of transport data sharing. It will also identify 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”;
- D2.9, D2.10 “Use cases definition v1 and v2”.

The overall WP3 goal is to analyse the market’s state-of the-art and enhance the potential impact of digitalisation and data sharing on different actors and on different areas of mobility and transport (economic, social, environmental, etc.). The following reports will be produced as deliverables of research data type:

- 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”;
- D3.5 “Societal and Environmental Impacts of data sharing assessment framework”.

These data are open to the public and are accessible via the repository chosen by the consortium.

## 2.3. Evaluation data

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. They include the KPIs measured for the data sharing culture, in order to determine a list of requirements that promote data sharing in the context of exploration and collaboration among actors and a projection of the data sharing ecosystem in transport and mobility for the coming years in terms of qualitative estimations.

First, WP3 aims to guide the evaluation of the new data sharing services and of new business models. In particular, the objective of Task 3.5 “Societal and Environmental Impacts of data sharing assessment framework” is 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 the society and on the environment. The task will propose a methodology to assess the relevant KPIs, by considering both quantitative and qualitative evaluations. A specific focus will be dedicated to creating a survey that will be used in Task 5.2 to evaluate:

- a) the acceptance of the Transport Cloud services;
- b) the business models for data sharing services. By month 15 of the project, this evaluation framework will be ready and validated by the General Assembly for further use in the Living Labs instances.

Next, in WP5, Task 5.2 “Quantification and measurement of the data exchange culture” will measure the impact of the data exchange culture by assessing the impact of Transport Cloud services on use cases challenges. It will also assess users’ acceptance of the Transport Cloud and of the business models. To that end, making use of the framework developed in T3.5, the assessment will evaluate, qualitatively but also quantitatively when possible, the hindrances and opportunities to data sharing 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 will evaluate other use cases as well, exploring and expanding on the recognised opportunities paving the way for new projects, creating requirements for further research but also providing input for improving the Transport Cloud’s functionalities. The output will be an analysis of the impact of the *Data Exchange Culture* by considering feedback from the participants to the Living Lab, that - during the project - actually use the Transport Cloud services and – after the project - might be potential uptakers of the business models. As a result, Task 5.2 will produce D5.2 “Report on Quantification and measurement of the data exchange culture” and D5.3 “Analysis and conclusions on the data exchange culture”.

These evaluation data will be accessible to the public via the repository of the consortium’s choice.

## 2.4. Technical data

The technical data category includes the data related to the technical developments of the MobiDataLab project, which correspond to the contents of WP4 and WP5.

WP4 will leverage on the open solutions identified in WP2 (Open Knowledge Base and use case requirements) in order to prototype a platform for searching, accessing and fusing multimodal mobility data in the cloud. WP4 aims to showcase the most effective tools to facilitate access to and exchange of mobility data (by humans and machines) for MaaS companies and developers. The resulting deliverables are:

- 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 will evolve towards actual and real-time data usage that will be fostered by physical Living Labs’ implementations in different locations and a Virtual Lab that will support the process and foster data exchange for the entire duration of MobiDataLab. The living labs in MobiDataLab serve two main purposes. First, to assess and measure, qualitatively and quantitatively, the data sharing culture and second, to generate new tools, technologies and insights through the utilisation of the MobiDataLab tools and data leading to an evaluation of the MobiDataLab’s 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 technical data mentioned above are public except D5.4, D5.6, D5.7 and D5.8 which are confidential and are available only for the members of the consortium.

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)<sup>1</sup>.

*Table 1 Dataset description template*

<b>Identifier of the dataset (URI)</b>	Identifies each dataset by a persistent URI <sup>2</sup> .
--	--

<sup>1</sup> <https://www.w3.org/TR/dwbp>

<sup>2</sup> Phil Archer; Nikos Loutas; Stijn Goedertier; Saky Kourtidis. Study On Persistent URIs. 17 December 2012. URL: <http://philarcher.org/diary/2013/uripersistence/>

<b>Name of the dataset</b>	Name of the dataset.
<b>Descriptive metadata</b>	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.)
<b>Data provenance</b>	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.
<b>Structural metadata</b>	Provides metadata that describes the schema and internal structure of a distribution, usually the properties or columns of the dataset schema.
<b>Data licenses</b>	A link to, or embedded copy of, a human-readable license agreement and/or machine-readable license agreement (e.g. ODBL, Open Data, etc.).
<b>Data formats and standards</b>	Machine-readable standardised data format under which the data is made available. Standardised data formats enable interoperability as well as future uses.
<b>Personal data (GDPR)</b>	Provides information about personal data and mention if the data is anonymised or not. Tell if the dataset entails personal data and how this issue is taken into account.
<b>Data preservation</b>	The preservation guarantee and the data storage during and after the project (for example databases, institutional repositories, public repositories, etc.).

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

An example description of a mobility-related dataset is available on the Data on the Web Best Practices published by the W3C<sup>3</sup>.

<sup>3</sup> <https://www.w3.org/TR/dwbp/dwbp-example.html#dataset-description>

## 3. Methodologies and Standards

### 3.1. FAIR principles

The FAIR principles refer to a concise, domain-independent, high-level and measurable set of guiding principles and practices to apply on a wide range of scientific data or metadata. They are the result of the work in 2014 of a community of stakeholders representing academia, industry, funding agencies, and scholarly publishers, which were then adopted the same year by the European Commission as the data guidelines for the Horizon 2020 (H2020) framework programme. They put “specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.”<sup>4</sup> The term “FAIR” refers to the characteristics of data or metadata of being Findable, Accessible, Interoperable and Reusable. 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, facilitate their support on 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.

These principles constitute then a general guide to FAIRness of data. But they are not themselves a standard or a specification. Precisely, they precede implementation choices and do not necessarily suggest any specific implementation solution. Instead, they act as a guide for data implementers, publishers and managers to evaluate whether their particular implementation choices are rendering their digital research artefacts FAIR. They form the basis for a long-term care of valuable digital assets composed by the data produced by the research project, while keeping the goal of being discovered and re-used by further research.

Given the definitions of FAIRness, there is however a clear distinction between FAIR data and Open data, the former does not necessarily imply the latter.<sup>5</sup> Indeed, while the openness of data is encouraged within H2020 programme (cf. 5.1.1 *infra*), 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 organisation, in order to make them more usable and valuable. Following the principle of “as open as possible, as closed as necessary”<sup>6</sup>, research data should be open by default, while setting a variable degree of openness. As illustrated by Figure 1, the more the data become both open and FAIR, the higher the benefits they bring.

---

<sup>4</sup> 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]

<sup>5</sup> [https://ec.europa.eu/info/sites/default/files/turning\\_fair\\_into\\_reality\\_0.pdf](https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_0.pdf)

<sup>6</sup> [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)

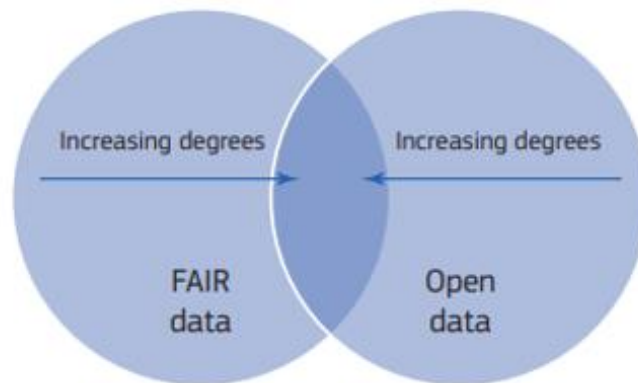


Figure 1 The relationship between FAIR and Open data

In general terms, the research data produced within H2020 should be FAIR. The MobiDataLab project intends to build 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 (REST API), a toolbox for dataset enrichment and quality raising, containing so-called processors (Semantic, Geographical, data format translations, injection of license specification and other enrichment processors). As specified in the Exploitation plan as deliverable D6.8 produced in T6.6, AKKA as the Exploitation manager of MobiDataLab, will be in charge of the Research Data Management, including the respect of application of FAIR principles and the constant update of referencing platforms like Zenodo<sup>7</sup>. The present DMP reflects also the decisions that affect the FAIRness of the data produced by MobiDataLab. After its first submission, there will be two other iterations to ensure its update for specification and documentation of the actions until the end of the project.

In the following sub-sections, the guiding principles of each one of the four concepts of FAIR are cited. We are also detailing how MobiDataLab will fulfil the requirements of FAIR principles.

### 3.1.1. Making data Findable

To be 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 (OKB), the Transport Cloud and the Labs (Living and Virtual). OKB will be produced as a result of WP2 and will provide a web-based tool for searching, querying a large set of resources linked to challenges and

<sup>7</sup> <https://zenodo.org/>

emerging principles for data sharing. Besides, via WP4, MobiDataLab aims at developing the Transport Cloud, a prototype of platform for searching, accessing and fusing multimodal mobility data in the cloud. This platform will concentrate the most effective tools to facilitate access to and exchange of mobility data (by humans and machines) for MaaS companies and developers. It will allow the discovery of data for operations or research, static and dynamic, real-time data and historical data, thus making data Findable.

### 3.1.2. *Making data Accessible*

To be Accessible

- A1. (meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorisation procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

In compliance with the H2020 rules regarding Open Access (OA) (cf. 5.1 *infra*) to scientific literature, MobiDataLab will make any scientific publication accessible online for free under the scope of the project. We will choose the most suitable approach (either “green” OA or “gold” OA) to peer-reviewed scientific publications that might result from the project. The publisher will be chosen amongst those who respect both the authors’ interests and accept the terms of open access publication (with an embargo period). On one hand, research data will benefit from an open access in a specific part of the project website, tailored to different levels of internal and external stakeholders. On the other hand, Partners will use an open access repository, connected to the tools proposed by the EC (e.g. OpenAIRE), in order to grant access to the publications and bibliographic metadata in a standard format, including the information requested by the EC.

### 3.1.3. *Making data Interoperable*

To be Interoperable

- I1. (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. 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 prototype platform produced in WP4 will:



- Ease the conversion from one standard to the other, e.g. NeTEx<sup>8</sup>, GTFS<sup>9</sup>, SIRI<sup>10</sup>, MDS<sup>11</sup>, etc.
- Trigger and provoke 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, completely documented and avoiding breaking changes) and able to deal with real-time data, providing data up to date / archived data, and making them Interoperable.

### 3.1.4. Making data Reusable

To be reusable

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 will prototype data processors in order to add value to the data. Precisely, Task 4.4 will start with the identification of datasets with missing values, which could greatly be enhanced by combining the data with other datasets and gathering additional results. Different data enrichment techniques will be provided as open tools.

More specifically, this task aims to:

- Enrich data semantically (combining with the Linked Open Data cloud, RDF/SPARQL);
- Enrich data geographically (tools for geocoding / projection, cross-referencing with spatial datasets, geodata APIs, OpenStreetMap data, etc.);
- Enrich data through complementary presentations (maps, tables, RDFs, etc.);
- Enrich data through machine learning;
- Verify that the enrichment did not distort the statistical power of the data.

To sum up, the transport cloud prototype produced in WP4 is primarily designed to demonstrate and offer solutions to reduce and, in some cases, remove current technical limitations identified as barriers to data reuse.

<sup>8</sup> <http://netex-cen.eu/>

<sup>9</sup> <https://gtfs.org/>

<sup>10</sup> <https://www.siri.org.uk/>

<sup>11</sup> <https://www.openmobilityfoundation.org/about-mds/>

## 3.2. GDPR

Within European Union, during the early stages of the internet, the 1995 Data Protection Directive was adopted. Over the last 25 years, tremendous changes in technology bring the need to revise data protection regulation. In 2016, the EU adopted the General Data Protection Regulation (GDPR). Now GDPR is recognised as 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 organisation (data collector, publisher, implementor) - regardless of geographical location - that collects, uses, or processes European Union citizens' personal information. Personal data security is thus improved via the enforcement of the four following aspects:

- Right to be forgotten: Article 17 mentions the obligation of the organisation, 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 25 stipulates that organisations 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 32 requires that organisations be able to prove an implementation of measures to ensure appropriate levels of security;
- Timely breach notifications: Article 34 imposes hefty fines to organisations 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 TRL. Amongst the legal factors, GDPR plays a data protection role and its "purpose limitation" principle (requirement that data collected for one specified, explicit, and legitimate purposes, be not processed for a new, incompatible purpose) can hinder our progress. For this reason, MobiDataLab will anticipate these limitations thanks to legal studies in WP2, and include an assessment of factors influencing data governance, including applicable legal regimes such as the GDPR.

### 3.2.1. *Personal Data Collection, Processing, and Re-use*

MobiDataLab will involve personal data collection and processing by a limited number of partners. It will also involve tracking or observation of participants arising from datasets and the further processing of previously collected personal data (i.e. secondary use). MobiDataLab will be subject to the provisions of the GDPR which enshrines the following key principles (without considering exemptions):

- Data must be processed fairly, lawfully and only for the purpose for which it was collected and further processed;
- Data cannot be disclosed without authorisation 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;
- Organisations 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;
- Organisation must have adequate security measures in place;
- The MobiDataLab consortium exercises a principle of minimum resort to notification exemptions afforded under national laws.

### 3.2.2. *Appointment of Data Protection Officer*

AKKA represented by Thierry Chevallier as Coordinator of MobiDataLab will act as the point of contact for Data Protection Issues and as the *de facto* Data Protection Officer (DPO) in the project.

The responsibilities of the DPO will be in line with Article 39 of the GDPR and will include, at a minimum, the following:

- Advise data controllers and processors within the MobiDataLab project on the processing of personal data, training of researchers and assignment of responsibilities;
- Provide support on the performance and tracking of Impact Assessments;
- Assist in risk assessment of personal data processing;
- Cooperate with any national or European supervisory authority and act as contact point for the project with such authorities.

### 3.2.3. *Compliance with GDPR Recital 78*

It is necessary for the data to be related to an identified or identifiable living individual. The individual need not be directly identifiable but may be identified by a reference number or some other tag which, in a given 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 anonymisation process(es) will be 'one-way,' with original source data being disposed of such that re-identification of data or decoding of anonymisation tokens by reference to any 'real-world' datasets will be rendered difficult to the greatest extent possible. MobiDataLab will follow the guidance set forth in the Article 29 Working Group 05/2014 Opinion on Anonymisation Techniques<sup>12</sup> and specifically its recommendations on Pseudo-anonymisation, Noise addition, Substitution, Aggregation, K-anonymity, L-diversity,

<sup>12</sup> [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf)

Differential privacy and Hashing/Tokenisation. MobiDataLab will also include downstream contractual obligations as a legal measure to respecting privacy in the use of the projects results.

### 3.2.4. *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. MobiDataLab is also likely to be exempted from national notification processes because our data collection is for the purposes of scientific research and because thus also additional institutional data protection measures, access restrictions, etc. will be put in place. We are mindful nevertheless that anonymisation 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. We also undertake to notify data protection authorities in jurisdictions where research activities will be carried out and specific relevant actions within it in order to obtain, if necessary, authorisation for such activities. The exact requirements and due diligence will need to be scoped and defined within the relevant jurisdictions.

The relevant national approvals will be 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 completion of all notification procedures.<sup>13</sup> Renewal of notifications, when necessary, will be carried out in-line with requirements of different national legislations. Processes for notification varies from one jurisdiction to another. The following project principals have been assigned the responsibility of acting as interlocutor with their own national data protection agency. Further information on notification procedures and the relevant agencies in Europe can be found in the Article 29 Working Group document “Vade mecum on Notification Requirements”

The exact partner and contact person who will notify the relevant Data Protection Agency will be determined during the project and after the transition from Directive 95/46 to General Data Protection Regulation (GDPR) 2016/679.

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

### 3.2.5. *Compliance with Article 49 of the GDPR*

While our position as scientific researchers permit us derogation from the prohibition on processing (sensitive categories of) Personal Data, we are nevertheless aware that it remains incumbent upon us, to provide specific and suitable safeguards so as to protect the fundamental rights and privacy

---

<sup>13</sup> <http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/others/2006-07-03-vademecum.doc>.

of Data Subjects. Some of these safeguards are already detailed above. MobiDataLab further undertakes to ensure that any Personal Data collected will also be treated in accordance with Article 49. In particular, Personal Data collected will be processed fairly and lawfully. Personal Data collected will be used only for research purposes as specified in our original proposal. The data will be adequate, relevant and not excessive in relation to the purposes for which they are collected. We will endeavour not to collect, and we will expunge all data that is not directly project related.

### *3.2.6. Safe Harbour and Privacy Shield Considerations*

In light of the Court of Justice of the European Union October 6th, 2015 decision on the EU-US Safe Harbour agreement, the MobiDataLab project will store all data derived from personal data (after anonymisation or dissociation) in EU member states and comply with the Article 29 Working Groups communiques on transfer of data outside of the union and forthcoming member state decisions on Safe Harbour.

### *3.2.7. 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 cognisant of concerns that arise in the Work Plan, MobiDataLab will take the following steps, in addition to those detailed above, toward addressing these:

- Compliance with the policy recommendations made in the Social Impact Expert Working Group EC DG ENTR Report (2012);
- The availability of 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 Design, 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;
- Test and evaluation of research results will be carried out in WP3 under the principal of informed consent.

### *3.2.8. 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 of certain categories of sensitive personal data (e.g. Article 6 and Recital 50). Exceptions for a right to opposition for processing or storage of data (Article 89), and for processing of data without consent (Article 6.1.f, Recitals 47 and 157) may be applicable. MobiDataLab commits to a minimum resort to exceptions and derogations in the processing of personal data within the project for the purposes of research.

### 3.2.9. *Activities Dedicated to Protection and Securing of Personal Data*

As project coordinator, AKKA shall ensure the consortium guarantees the treatment of personal data generated during the project. This will be 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 will also integrate the technical requirements of European, national and regional data protection legislation. Partners will be 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 will be 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 purpose;
- Policies for Data Controller and Data Processors;
- Description of the technical characteristics of the data processed;
- Technical features and topology of the information systems where data is stored and processed.

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

- The Charter of Fundamental Rights of the EU (especially articles 3, 7, 8 and 25);
- Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use;
- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data;
- Convention for the Protection of Individuals with regard to automatic processing of personal data (Convention Nr. 108);
- WMA Declaration of Helsinki (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;
- CIOMS -- International Ethical Guidelines for Biomedical Research Involving Human Subjects.
- Code of Nuremberg (Article 10);
- Ethics and EU funded research Council Decision 1513/2002/EC on FP6 (e.g. article 3).

### 3.2.10. Shared Information and Personal Data

The partners agree that any background, results, confidential information and/or any and 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 Article 2, Section (a) of the Data Protection Directive (95/46/EEC) (hereinafter referred to as “Personal Data”) or under Article 4.1 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. de-identify) to the other partners prior to providing the shared information to such other partners.

## 4. Allocation of resources

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

All research data collected as part of this project is owned by the data producer or partners involved in trial sites. The partners in MobiDataLab will take the responsibility for the collection, management, and sharing of the research data. Quality assessment will be the responsibility of data manager of each trial site.

The costs to make the data FAIR in MobiDataLab shall be handled by each partner who will have to generate its data according to the requirements expressed in the Data Management Plan. Specifically, different tasks of WP4 will contribute to fulfil the requirements of 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, responsible, spatial coverage, media type, etc.) allowing both human understanding and automatic discovery by software agents, thus making data *Findable*. This task is led by AKKA with a resource-effort of 6 person-month (PM) distributed from month 6 to month 32 of the project.

Second, Task 4.3 Data Access Services and Data Channels performs the practical integration of available data access services and data channels. Availability, documentation and usability will be 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 will be connected and become *Accessible*. The activities are connected to parallel tasks in this context and will generate a positive impact. HERE is the Task Leader and has a planned effort of 3 PM for the period from month 6 to month 32 of the project.

Third, MobiDataLab data will be made *Interoperable* by two tasks. On one hand, via Task 4.3 Data Access Services and Data Channels, the Transport Cloud prototype will provide 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 contributes to the development of open tools allowing the enrichment of data. For example, tools for spatial enrichment will include geocoders to convert a human readable address into spatial coordinates, geo converters to simplify, convert or normalise geographical data (e.g. projections), spatial functions to run computations (e.g. distance computation) on them. Such data format translations make data interoperable. AKKA is the Leader of Task 4.4 which manages 8 PM on this task from month 6 to month 32 of the project.

Fourth, Task 4.4 Data Processors starts with the identification of datasets with missing values, which could greatly be enhanced by combining the data with other datasets and gathering additional results. Different data enrichment techniques, including semantic enrichment, geographic transformations, will be provided as open tools, generating new data and encouraging *Reuse* by adding value. Moreover, Task 4.5 Anonymisation and Data Privacy includes data processing modules that apply data protection and anonymisation techniques to the mobility data in the Transport Cloud. New protection techniques will be researched and developed where current state of the art techniques cannot fulfil the project requirements, thus promoting high added value data reuse. Task 4.5 Leader is URV which has at their disposal 18 PM throughout the task period running from month 6 to month 32 of the project.

## 5. Access and Curation

### 5.1. Data Access

#### 5.1.1. *Open Access and ORDP*

Within Horizon 2020, an open access to scientific information is an objective. The notion of open access here refers to “the practice of providing online access to scientific information that is free of charge to the end-user and reusable”<sup>14</sup>, where “scientific information” includes both the peer-reviewed scientific research articles/publications and the research data underlying publications, curated data and/or raw data.

To ensure open access to all peer-reviewed scientific research publications relating to project results, which is an obligation, two routes are possible:

- Self-archiving / 'green' open access - the author, or a representative, archives (deposits) machine-readable electronic copy of the published version or the final peer-reviewed

<sup>14</sup> [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)



manuscript in an online repository before, upon or after its publication. Some repository software delays access only after an embargo period has elapsed. If this is the case, the European Commission demands that the open access is ensured within a maximum of six months;

- Open access publishing / 'gold' open access - an article is immediately published at a publisher or on a journal website. 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.

Whichever the route chosen, H2020 beneficiaries must at least ensure that any scientific peer-reviewed publications can be read online, downloaded and printed. They must also strive to provide the right to copy, distribute, search, link, crawl and mine to the public, in order to make publications more useful. As mentioned previously (cf. 3.1.2 *supra*), peer-reviewed scientific publications resulting from the project will become accessible openly, thanks to an open access repository used by partners, connected to the tools proposed by the EC (e.g. openAIRE), which grants access to the publications and bibliographic metadata in a standard format, including the information requested by the EC.

With regard to the openness of research data, the Open Research Data Pilot (ORDP) run by the European Commission “aims to improve and maximise access to and re-use of research data generated by H2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.” This ORDP applies primarily to the digital form data needed to validate the results presented in scientific publications, including statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge.

Despite the benefits of open access, some research data cannot be made open. For this reason, the principle of “as open as possible, as closed as necessary” applies. It is therefore “possible to opt out of research data sharing at any stage - before or after the signature of the grant agreement - but reasons have to be given e.g. for IPR concerns, privacy/data protection concerns, national security concern, if it would run against the main objective of the project or for other legitimate reasons.”, precisely the same potential reasons which may play a role in the balance.

In order to specify the type of data subject to open access, a Data Management Plan is required, which is the subject of this present deliverable. The management of scientific information within H2020 under the open access guidelines is illustrated as follows:

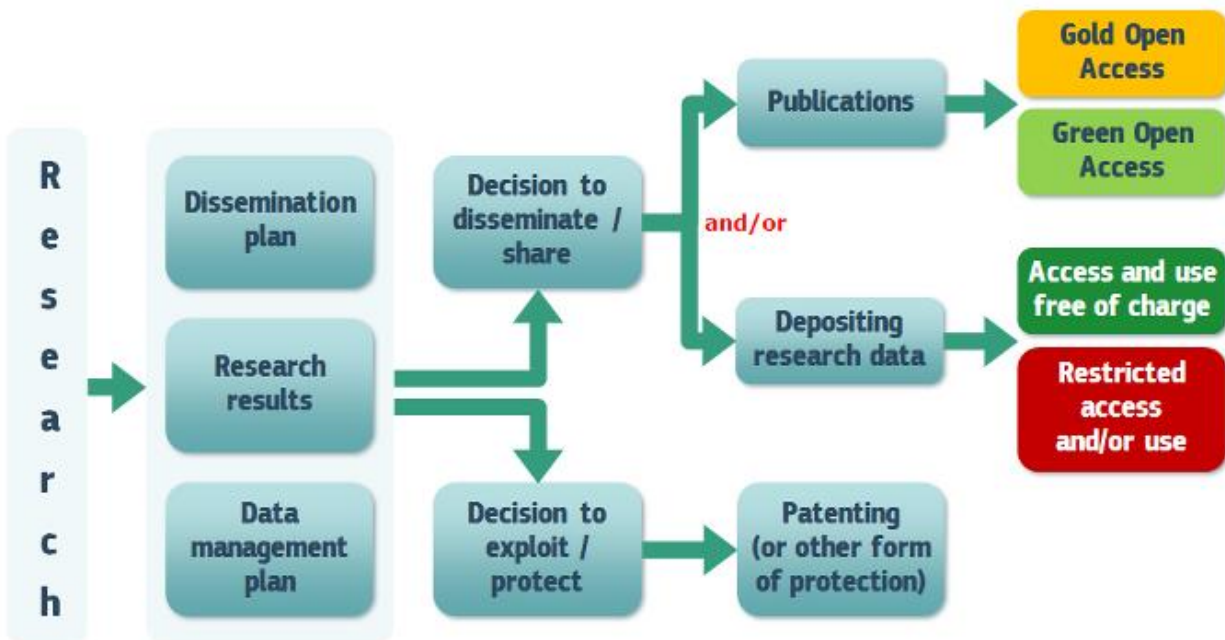


Figure 2 Open access to research data

As mentioned in the Grant Agreement, MobiDataLab is fully committed to ORDP.

As already explained previously, the two pillars to ORDP consist of “the development of a Data Management Plan and the open access to research data, if possible”<sup>15</sup>. The first point is being addressed by the establishment of a DMP as a deliverable of MobiDataLab and by its maintenance. The second point will be fulfilled by the open access to research data including public deliverables, non-protected results or results coming after a patent registration, which will be rendered throughout the whole project’s duration in a specific part of the project website, tailored to different levels of internal and external stakeholders but also published on a general purpose open repository like Zenodo. The partners will agree altogether on the data that are shareable and the ones that are confidential and that should remain in the consortium only.

### 5.1.2. Data Licensing

The results of the MobiDataLab project, i.e. data, knowledge or information produced and generated, are attached to the intellectual property rights to use them. In absence of specification, national IPR apply. However, national IPR are different from one country to another, thus it is difficult to interpret voluminous national laws of several countries. Short terms of use adapted for the special case are usually the better solution.

<sup>15</sup> <https://www.openaire.eu/what-is-the-open-research-data-pilot>

In order to increase data reuse on the results of the project, partners should assign a license to their shared data, so that results can be exploited. A license describes the conditions under which the data or software is (re)usable. The license should contain:

- a disclaimer of warranty and limitation of liability;
- a citation requirement;
- rules for databases and collections.

For example, Creative Commons Corporation (CC) has introduced a well-known system of licenses.<sup>16</sup> CC licenses are free of charge and easy to use since any license consists of a legally binding arrangement, a summary, a symbol, which may be downloaded as graphics file and inserted in a document, and a machine-readable symbol. Their public copyright licenses incorporate a unique and innovative “three-layer” design: Legal Code, Commons Deed and CC REL. This type of licence may be utilised for protection of all types of content. It also makes the concept understandable by the creators of works, uses and the Web itself.

Besides, Open Source Initiative (OSI) approves some licenses which comply with the Open Source Definition and which have gone through the OSI license review process. Such approved licenses allow software to be freely used, modified, and shared.<sup>17</sup> They include Apache, GNU, MIT, Mozilla, just to mention the most popular ones. The consortium will choose the most appropriate licenses for use depending on the type of data concerned.

### 5.1.3. Storage

In the European framework, European Open Science Cloud (EOSC) is an initiative providing a world-class data infrastructure to store and manage data. It arises from the EC’s aim to promote the access and reuse of research data which comes out of publicly funded research. Under H2020, EOSC becomes the best instrument to provide a framework for collaboration and the pooling of resources at European, national, regional and institutional levels.<sup>18</sup> In particular, EOSC aims at solving 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 of open access or even fully open access. For example, BlueBRIDGE suggests an online environment to support secure and controlled data sharing and storage, INRIA’s Software Heritage archive is a great library of source code on engineering and technology sciences, and Institute of Informatics of Slovak Academy of Sciences offers EGI FedCloud client.

EOSC provides also other services such as Open Science publishing infrastructure which is connected to the EOSC. OpenAIRE is one of such examples which contributes actively to EOSC.<sup>19</sup>

<sup>16</sup> <https://creativecommons.org/licenses/>

<sup>17</sup> <https://opensource.org/licenses>

<sup>18</sup> <https://eosc-portal.eu/about/eosc>

<sup>19</sup> <https://www.openaire.eu/openaire-and-eosc>

On one hand, OpenAIRE provides a cost-free data repository Zenodo allowing data deposit and discovery, rendering data Accessible. On the other hand, through harvesting repositories and mining techniques OpenAIRE infers links between publications, research funding and research data, which enables data Reuse.

As mentioned in the previous section concerning Open access, MobiDataLab partners will use such an open access repository, as OpenAIRE, in order to grant access to the publications and bibliographic metadata in a standard format. Specifically, Zenodo is a repository where every upload is assigned a Digital Object Identifier (DOI), making them citable, trackable, thus Findable. It also provides flexible Access conditions. Zenodo also ensure Interoperability by allowing GitHub integration to preserve metadata. Finally, its versioning gives the possibility to update datasets easily and store all changes in metadata over record's lifetime, which enable Reuse. To summarise, these advantages are totally compliant with the FAIR principles.

### 5.1.4. 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 data servers should be authenticated. Also, proper means are used to authenticate the servers. An authentication system could be used to handle the authentication of the users during the project;
- Authorisation: The access to MobiDataLab data servers is only available to authenticated and authorised 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 each trial site and project wide to provide the authorisation;
- Accounting: In MobiDataLab, any access and modification to a resource by any user is securely logged in order 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 servers should be encrypted during transmission and storage;
- Communication Security: Access to MobiDataLab servers should be done through encrypted communication channels such as HTTPS and 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;
- Availability: This security principle assures that the MobiDataLab servers 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 DoS attacks should be implemented.

## 5.2. Data qualification and curation

### 5.2.1. Data protection

Securing stored digital data involves preventing unauthorised 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 taking action as needed. This means that a multi-tier approach needs to be adopted from 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 are a centralised theme in many high-profile breaches. Partners within MobiDataLab will follow a specific set of guidelines to comply with the project's main requirement for 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.

### 5.2.2. *Data qualification*

The aim of the data qualification is to clarify and thoroughly document the semantic of the data. This facilitates subsequent data analysis. If there is an original data documentation, it typically focuses on what is of interest to the management or the auditors. Yet, data qualification allows to explore the semantics of the data in details, which then updates or specify this documentation. Indeed, data qualification allows to understand and reveal the gap between the perceived semantic of current data analysers and the true and underlying semantic of the data.

Data are dynamic and require constant maintenance. Data qualification should thus be constantly carried out, so as to keep data up-to-date, valid, accurate and reliable. In order to operate data qualification, additional computation is needed to implement rules for different processes of data verification, validation, approvals and rejection. This generates metadata. Although qualifying data necessitates additional resources and efforts, better qualified data deliver higher quality and consistency which help to make decisions and clarify the direction of further research.

## 6. Conclusion

This Data Management Plan V1 gives a first overview of the data processed in MobiDataLab and describes the data categories of the project, providing information on their management and on the

implementation of the FAIR principles in the project. MobiDataLab will be committed to the principles of the H2020 Open Research Data Pilot, and therefore the guidelines associated with open access have been described, with the aim to ensure that the project's outputs will be openly available to the research community.

*Towards the Data Management Plan V2:* It is important to note that the Data Management Plan is a living document and will be constantly updated until the end of the MobiDataLab project. The present document is the first version of the Data Management Plan, and in the next version (D1.5, 2nd issue), more details will be provided, especially regarding the description of shared datasets, and corresponding standards and methodologies. A great consideration will be given to mobility datasets to be shared in the course of the Living and Virtual Labs (WP5) and for which we do not yet have sufficient information at this stage of the project. Indeed, MobiDataLab being a project about mobility data sharing, a particular attention obviously needs to be paid to mobility datasets. Mobility data categories will be defined, relying on standardisation works from initiatives like Transmodel, for example, by making distinctions between transport offer data (whether they are dynamic, i.e. real-time, or static) and customer usage data. These different categories of data require different treatments and the definition by the project of an appropriate methodology (for example anonymisation of personal traveller data). In the same way, the data from our reference group of stakeholders (local authorities, public transport operators, stakeholders, etc.) will be detailed, and their specificities duly described.

## MobiDataLab consortium

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



[@MobiDataLab](https://twitter.com/MobiDataLab)  
#MobiDataLab



<https://www.linkedin.com/company/mobidatalab>

For further information please visit [www.mobidatalab.eu](http://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.