MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D1.5 Data Management Plan (2nd issue)

11/08/2022
Authors: Samantha DRAY (AKKA)

Summary sheet

Deliverable Number	D1.5
Deliverable Name	Data Management Plan (2nd issue)
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Samantha DRAY (AKKA)
Contributing Partner(s)	AETHON, AKKA, CNR, F6S, HERE, ICOOR, KISIO, KUL, POLIS, URV
Peer Review	HERE, URV
Contractual Delivery Date	31-07-2022
Actual Delivery Date	28-07-2022
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	39
WP/Task related to the deliverable	WP1/T1.4
WP/Task responsible	AKKA/AKKA
Document ID	MobiDataLab-D1.5-DataManagementPlanV2-v1.0
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

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 [formerly: KISIO DIGITAL]	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	31/03/2022	AKKA	Table of content	Initial version
0.2	10/07/2022	AKKA	All	Draft version
0.3	24/07/2022	HERE, URV	All	Peer Review
0.4	26/07/2022	AKKA	All	Rework
0.5	26-28/07/2022	AKKA	All	Quality Check
1.0	28/07/2022	AKKA	All	Submission

Executive Summary

This report is the second 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. This initial version 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 (this deliverable D1.5) is implemented as a "living" DMP that is continuously updated throughout the course of the project. It reflects the status of the collected, processed or generated data and follows what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

Table of contents

List of figures

List of tables

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[1]) 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) 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) is providing 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] OpenAIRE – How to Make Your Data FAIR?

MOBIDATALAB

1.3. Intended audience

The dissemination level of this D1.5 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 relationship 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 the 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 that 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 into 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 for historical purposes and follow-up within WP1 such as meeting minutes, recordings, internal reports…

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 to access all available functionalities linked to the project. These administrative data are confidential and are available only to the members of the consortium.

On the other hand, 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: MobiDataLab.

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 different areas of mobility and transport (economic, social, environmental, etc.). The following reports will be produced as deliverables of the 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, 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 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 society and 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 the business models. To that end, making use of the framework developed in T3.5, the assessment will evaluate, qualitatively but also quantitively 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 and 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 the open solutions identified in WP2 (Open Knowledge Base and use case requirements) 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 for 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)[2].

Table 1 Dataset description template

Identifier of the dataset (URI)	Identifies each dataset by a persistent URI[3].
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

[2] https://www.w3.org/TR/dwbp
[3] Phil Archer; Nikos Loutas; Stijn Goedertier; Saky Kourtidis. Study On Persistent URIs. 17 December 2012. URL: http://philarcher.org/diary/2013/uripersistence/

	describe provenance information, such as W3C's Provenance Ontology.
Structural metadata	Provides metadata that describes the schema and 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 mention 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 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 in the "Data on the Web Best Practices" published by the W3C[4].

---

[4] https://www.w3.org/TR/dwbp/dwbp-example.html#dataset-description

MOBIDATALAB

Funded by the European Union

3. Data Management in MobiDataLab

3.1. About Data management in MobiDataLab

MobiDataLab as an H2020 project is obliged to provide a continuously updated 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 should be OpenAIRE-compliant to enable the harvesting of metadata.
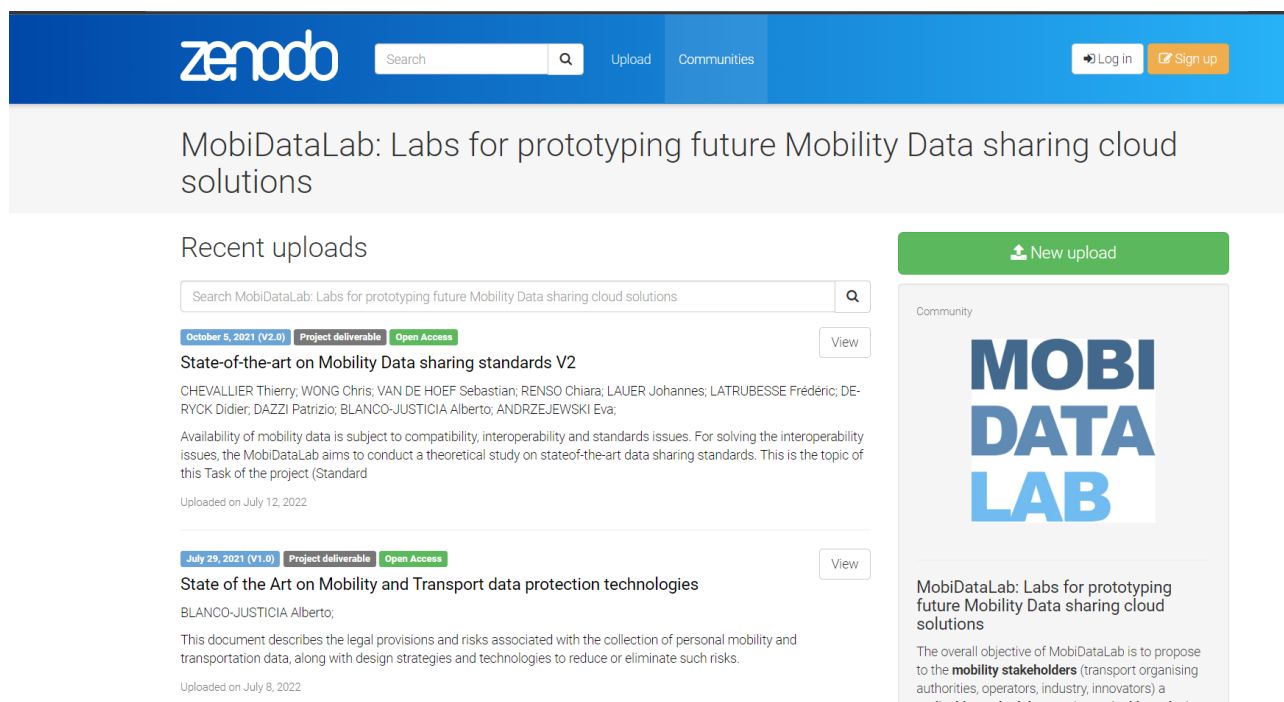

Figure 1 - MobiDataLab Zenodo Repository

A project-wide data management policy that handles these issues on the administrative and technical levels has been established and described in MobiDataLab deliverable D1.4 Data Management Plan v1 [1]. This includes for example topics like data and meta-data collection, personal data treatment, the data repository infrastructure and the mandatory compliance to OpenAIRE - the Open Access Infrastructure for Research in Europe (Figure 2).

Figure 2 - MobiDataLab Project in OpenAIRE

3.2. Roles & responsabilities

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's 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 act 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 for using MobiDataLab's living DMP
- 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 MobiDataLab's living DMP
- Offering customized help and further guidance for publishing open data and open-source software
- Monitoring 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 a publication of the data
- Ensuring that the metadata of data used and produced at the Work Package level is made available in MobiDataLab's living DMP according to the MobiDataLab data management policy and guidelines promptly.

AKKA represented by Thierry Chevallier as Coordinator of MobiDataLab acts as the point of contact for Data Protection Issues and is accompanied by Samantha Dray acting as Data Protection Officer (DPO) in the project.

WP Data Managers are named inside each Work Package.

3.3. The MobiDataLab data management

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 & Work Packages leaders.

Figure 3 - MobiDataLab page on DMPOnline

3.4. The Transport Cloud & The Open Knowledge Base (OKB).

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 for adding value to the data
- demonstrating anonymisation and privacy-preserving tools

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

MOBIDATALAB
MOBIDATALAB – H2020 G.A. No. 101006879

Funded by the European Union

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 will use CKAN, GeoNetwork and OpenDataSoft as Data Catalog demonstrators.

All information related can be found _inside_ D4.1 Transport Cloud Architecture Dossier and will be available in D4.3 Reference Data Catalog.

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 will be available to the public by the first trimester of 2023 and will be 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 used by MobiDataLab

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

The respective datasets will be in a Transport Cloud using CKAN/GeoNetwork/OpenDataSoft and could be assigned the tags **input-data** and **open data** (in addition to the respective DC and WP tags) and added to the group "Open data used by MobiDataLab".

The final version of the DMP will provide an exhaustive list of the data sources.

Such datasets may encompass datasets that are collected from public authorities and institutions.

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

The MobiDataLab's Transport Cloud Data Catalogue will provide some meta-data related to the usage of the data in MobiDataLab.
The actual data can also, in general, be downloaded directly from the websites of respective organisations 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 visualisation, the respective CKAN resource (data) associated with the CKAN dataset (meta-data) links to the original source of the dataset.
Thanks to the Transport Cloud Data Catalogue, companies or actors of public transportation can easily access meta-data of interest, therefore, gaining insights into their business.

3.6. Non-open Data used by MobiDataLab

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 DC and WP tags) and associated with a group "non-open data used by the MobiDataLab".

3.7. Open Data Created by MobiDataLab

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

The respective datasets will be either put on Zenodo and linked to the OpenAire project page and will also be accessible on the Open Knowledge Base for the documentation part and the Transport cloud for more Technical Data.

3.8. Non-open Data created by MobiDataLab

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 DC and WP tags) and associated with a group "non-open data used by the MobiDataLab"

4. Methodologies and Standards

4.1. FAIR principles

The FAIR principles refer to a concise, domain-independent, high-level and measurable set of guiding principles and practices to apply to 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."[5] 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, 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.

These principles constitute then a general guide to the 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 of a 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 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.[6] Indeed, while the openness of data is encouraged within the 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, to make them more usable and valuable. Following the principle of "as open as possible, as closed as necessary"[7], 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.

---

[5] 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]
[6] https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_0.pdf
[7] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

MOBIDATALAB
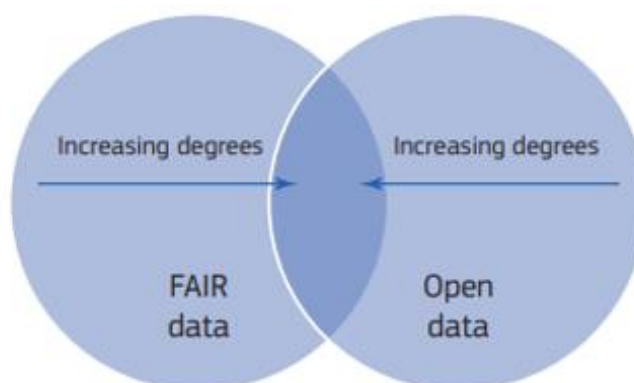MOBIDATALAB – H2020 G.A. No. 101006879

Figure 4 - 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 oversee the Research Data Management, including the respect of the application of FAIR principles and the constant update of referencing platforms like Zenodo[8]. The DMP also reflects the decisions that affect the FAIRness of the data produced by MobiDataLab. We are currently at the second iteration where the specification and documentation of the actions are being updated. There will be one more iteration at 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.

4.1.1. 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

---

[8] https://zenodo.org/

MOBIDATALAB

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 by querying a large set of resources linked to challenges and emerging principles for data sharing.

Besides, via WP4, MobiDataLab aims at developing the Transport Cloud, a prototype of a platform for searching, accessing and fusing multimodal mobility data in the cloud. This platform will concentrate on 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.

4.1.2.   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 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), to grant access to the publications and bibliographic metadata in a standard format, including the information requested by the EC.

4.1.3.   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[9], GTFS[10], SIRI[11], MDS[12], 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.

4.1.4. 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 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[13]/SPARQL[14])
- 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.

---

[9] http://netex-cen.eu/
[10] https://gtfs.org/
[11] https://www.siri.org.uk/ -- restricted access
[12] https://www.openmobilityfoundation.org/about-mds/
[13] https://www.w3.org/RDF/
[14] https://www.w3.org/2001/sw/wiki/SPARQL

MOBIDATALAB

Funded by the European Union

4.2. 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 regulations. 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 purpose, 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.

4.2.1. 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 the 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

MOBIDATALAB

- 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 the minimum resort to notification exemptions afforded under national laws.

4.2.2. Compliance with GDPR Recital 78

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 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 outlined in Article 29 Working Group 05/2014 Opinion on Anonymisation Techniques[15] and specifically its recommendations on Pseudo-anonymisation, Noise addition, Substitution, Aggregation, K-anonymity, L-diversity, Differential privacy and Hashing/Tokenisation. MobiDataLab will also include downstream contractual obligations as a legal measure to respect privacy in the use of the project's results.

4.2.3. 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. MobiDataLab is also likely to be exempted from national notification processes because our data collection is for 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

---

[15] http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

MOBIDATALAB

MOBIDATALAB – H2020 G.A. No. 101006879

within it 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 the completion of all notification procedures.[16] Renewal of notifications, when necessary, will be 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 own 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"

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 the 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.

4.2.4. Compliance with Article 49 of the GDPR

While our position as scientific researchers permits 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 to protect the fundamental rights and privacy 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 under 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 concerning 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.

4.2.5. Safe Harbour and Privacy Shield Considerations

Considering 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 Article 29 Working Groups communique on the transfer of data outside of the union and forthcoming member state decisions on Safe Harbour.

---

[16] http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/others/2006-07-03-vademecum.doc.

MOBIDATALAB

4.2.6. 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 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 principle of informed consent

4.2.7. 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 6 and Recital 50). 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 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 research purposes.

4.2.8. 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

MOBIDATALAB

Funded by the European Union

- 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 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 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 concerning the processing of personal data and on the free movement of such data
- Convention for the Protection of Individuals about 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)

4.2.9.   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 before providing the shared information to such other partners.

5. Allocation of ressources

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 of 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 the data controller 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 fulfilling 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) and 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 and geographic transformations, will be provided as open tools, generating new data and encouraging *Reuse* by

MOBIDATALAB

adding value. Moreover, Task 4.5 Anonymisation and Data Privacy include data processing modules that apply for data protection and anonymisation techniques to the mobility data in the Transport Cloud. New protection techniques will be researched and developed where the 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 its disposal 18 PM throughout the task period running from month 6 to month 32 of the project.

6. Access and Curation

6.1. Data Access

6.1.1.   Open Access and ORDP

Within Horizon 2020, 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"[17], 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) a machine-readable electronic copy of the published version or the final peer-reviewed 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 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 route is 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, 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.

Regarding 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

---

[17] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

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 concerns, 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.

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:
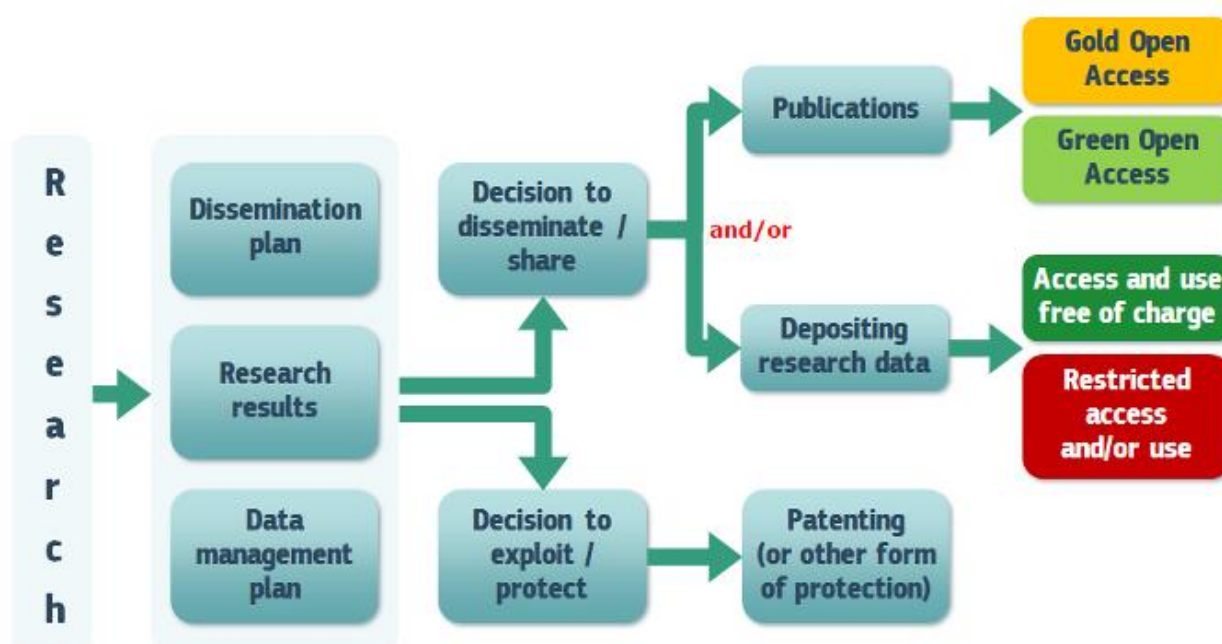

Figure 5 - Open access to research data

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

As already explained previously, the two pillars of ORDP consist of "the development of a Data Management Plan and the open access to research data, if possible"[18]. 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,

---

[18] https://www.openaire.eu/what-is-the-open-research-data-pilot

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 shareable data and the ones that are confidential and that should remain in the consortium only.

6.1.2. 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, nationals IPR are different from one country to another, thus it is difficult to interpret the voluminous national laws of several countries. Short terms of use adapted for the special case are usually the better solution.

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.[19] 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 files 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 the 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 the software to be freely used, modified, and shared.[20] They include Apache, GNU, Mand IT, and 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.

---

[19] https://creativecommons.org/licenses/
[20] https://opensource.org/licenses

MOBIDATALAB
MOBIDATALAB – H2020 G.A. No. 101006879

6.1.3.   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.[21] 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 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 the 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 example which contributes actively to EOSC.[22] 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, 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, and thus Findable. It also provides flexible Access conditions. Zenodo also ensures 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 the record's lifetime, which enable Reuse. To summarise, these advantages are totally compliant with the FAIR principles.

6.1.4.   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

---

MOBIDATALAB

MOBIDATALAB – H2020 G.A. No. 101006879

- **Authorisation**: 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 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

6.2. Data qualification and curation

6.2.1. 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 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 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 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

6.2.2. Data qualification

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

Data are dynamic and require constant maintenance. Data qualification should thus be constantly carried out, to keep data up-to-date, valid, accurate and reliable. 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.

7. Conclusion

First, 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 former Data Management Plan V1 gave a first overview of the data processed in MobiDataLab and described the data categories of the project, providing information on their management and the implementation of the FAIR principles in the project. MobiDataLab is 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.

The present document is the second version of the Data Management Plan, and in the next version (D1.6, 3rd issue), more details will be provided, especially regarding the description of shared datasets, and corresponding standards and methodologies. Great consideration is given to mobility datasets to be shared during the Living and Virtual Labs (WP5) and for which we do not yet have subsequent information at this stage of the project. Indeed, MobiDataLab being a project about mobility data sharing, particular attention obviously needs to be paid to mobility datasets. Mobility data categories are under the definition, 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.) are going to 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.

Two changes will occur in the next months in the MobiDataLab Consortium:

- One on name: KISIO becoming HOVE
- One on the entity: By fusion with the Addeco and Modis Group, AKKA will become AKKODIS

This will induce an updated version of this DMP to reflect those modifications.

[@MobiDataLab](#)
#MobiDataLab

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

For further information please visit **www.mobidatalab.eu**

Funded by the European Union