Fleet and traffic management systems
for conducting future cooperative mobility

D7.2 First version of Data Management Plan

Document Type	Deliverable
Document Number	D7.2
Primary Author(s)	Gregor Papa | JSI
Document Version / Status	1.0 | Final
Distribution Level	PU (public)
Project Acronym	CONDUCTOR
Project Title	Fleet and traffic management system for conducting cooperative mobility
Project Website	https://conductor-project.eu/
Project Coordinator	Netcompany-Intrasoft SA | www.netcompany-intrasoft.com
Grant Agreement Number	101077049

CONDUCTOR project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101077049.

EXECUTIVE SUMMARY

The Data Management Plan (DMP) is a central document that addresses all open internal research data management activities within the CONDUCTOR project. It provides a description of how the data generated, collected, and processed within the project will be managed. This is a working document that presents the first version of the DMP and complies with the Horizon Europe template for DMP. It outlines data management strategies to meet Horizon Europe research data management requirements.

The aim of CONDUCTOR DMP is to improve access and reuse of research data funded by the EU project and to support dissemination of generated knowledge to other European research organizations. Considering the balance between openness and protection of research data, the project consortium will follow the principle of "as open as possible, as closed as necessary".

The purpose of a DMP is to provide an analysis of the main elements of the project's data management policy, taking into account all data sets to be generated. Research data are defined as facts and numbers that are collected and observed, with the focus on digital data. The project consortium is aware that digital data must be handled with care. Therefore, the continued use of the data created in this project will be ensured through holistic data management. The DMP will help manage the generated research data.

In addition to providing information about the data generated in this project, the DMP outlines how curation, preservation, and sustainability of the data will be ensured. It also describes what parts of the research data will be open and the extent to which they will be available. In addition, the DMP shows the progress of the data during the project and specifies how the research data will be handled beyond the duration of the project. It also determines the accessibility of the data for reuse for own purposes and for use by third parties.

CONTRIBUTORS

Name	Organization	Name	Organization
Gregor Papa	JSI	Konstantinos Gkiotsalitis	NTUA
Miha Cimperman	JSI	Emmanouil Nisyrios	NTUA
Raquel Sánchez	Nommon	Oskar Eikenbroek	UTWENTE
Oliva García Cantú	Nommon		
Paola Lanzi	DeepBlue		
Francois Brambati	DeepBlue		

FORMAL REVIEWERS

Name	Organization	Date
Fynn Wolf	TUM	2023-04-21
Joan Estrada	BAX	2023-04-24

DOCUMENT HISTORY

Revision	Date	Author / Organization	Description
0.1	2022-12-15	Gregor Papa / JSI	Template setup and initial text and table of contents
0.2	2023-04-17	Gregor Papa / JSI	Populated document
0.3	2023-04-18	Gregor Papa / JSI	Document for review
0.4	2023-04-19	Gregor Papa / JSI	Updated document for review
0.5	2023-04-24	Gregor Papa / JSI	Review updates
0.6	2023-04-24	Gregor Papa / JSI	Final draft
1.0	2023-04-28	Flavien Massi / INTRA	Final check and submitted version

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

1.1 Project information

The project ***Fleet and traffic management systems for conducting future cooperative mobility*** **(CONDUCTOR)** will design, integrate and demonstrate advanced, high-level traffic and fleet management that will allow efficient and globally optimal transport of passengers and goods, while ensuring seamless multimodality and interoperability. Using innovative dynamic balancing and priority-based management of vehicles (automated and conventional) CONDUCTOR will build on state-of-the-art fleet and traffic management solutions in the CCAM ecosystem and develop next generation simulation models and tools enabled by machine learning and data fusion, enhancing the capabilities of transport authorities and operators, allowing them to become “conductors” of future mobility networks.

The project will focus on upgrading existing tools based on driver-centred approaches towards a mobility user-centred approach and ensuring that the objectives are realistically achievable, based on close collaboration with relevant stakeholders at operator/city/authority level, where user needs are identified and validated. Therefore, the CONDUCTOR project is built around the following objectives:

- To demonstrate traffic and fleet management to integrate CCAM for people and goods;
- To address intermodal interfaces and interoperability between traffic management systems;
- To test and demonstrate advanced simulation models in real-life traffic conditions considering different priorities;
- To demonstrate optimised mobility network load balancing;
- To consider governance of the traffic management system considering user needs.

The results of the integration of the models will be validated against three use cases, each of which covers some of the key call requirements for the interoperability of traffic management systems and the integration of different transport modes, taking into account people and goods:

- UC1 Integrated traffic management with intermodality: demonstrated in three cities – Athens, Madrid and Almelo.
- UC2 Demand-responsive transport: demonstrated in Slovenia.
- UC3 Urban logistics: demonstrated in Madrid.

1.2 Document Scope

This deliverable describes the Data Management Plan (DMP) and provides a description of how the data generated, collected and processed within the project will be managed. It is a working document that presents the first version of the DMP and complies with the Horizon Europe template for DMP. It outlines data management strategies to meet the research data management requirements in Horizon Europe. The purpose of a DMP is to provide an analysis of the main elements of the project's data management policy, taking into account all data sets that will be generated. Research data are defined as facts and numbers that are collected and observed, with a focus on digital data. The project consortium is aware that digital data must be handled with care. Therefore, the continued use of the data created in this project will be ensured through holistic data management. The DMP will help manage the generated research data.

1.3 Document structure

The structure of this document is as follows.

- Section 1 contains a brief introduction to the project.
- Section 2 provides a summary of the re-used and generated data.
- Section 3 describes the FAIR data approaches in terms of findability, accessibility, interoperability and reuse.
- Section 4 outlines some other research outputs.
- Section 5 gives an insight into data management resources.
- Section 6 is dedicated to data security.
- Section 7 deals with ethical aspects.
- Section 8 covers other related issues.
- Section 9 provides some concluding remarks.

2 DATA SUMMARY

As part of the project activities, CONDUCTOR will collect and manage different types of data. Datasets will be collected through desk research, interviews, workshops, simulations and use case studies. These datasets are collections of standard material produced as part of a research project, e.g., project deliverables and dissemination material.

At the beginning of the project, the CONDUCTOR partners contributed with existing traffic-related data (open data, simulations and other data collected in previous projects). These data are of different types and formats, which will be clearly specified in the next version of this deliverable, as shown in Table 1. The data sources belong to different categories, such as i) Transport supply data (road network data, public transport (PT) offer, taxi supply, etc.); ii) Transport demand data (delivery demand, traffic and passenger counts, ticket sales, taxi demand, etc.); iii) Socio-demographic data (census data, land use, etc.); iv) Survey data (demand estimates, declared travel preferences, etc.); v) Other data (mobile phone records/telecommunications data, weather data, traffic data, other monitoring data, etc.).

Table 1: Available existing data

Pilot	Tentative data	Accessibility
Athens Pilot	Public transport routes and stations (bus & metro)	NTUA, OASA & LIT
	Passenger data: data from ticket validator for busses & metro (using for optimization of scheduling)	NTUA, OASA & LIT
	OASA data: traffic events data (road blocks, etc.) - planned to use for Bus scheduling	NTUA, OASA & LIT
	Telematics of bus & metro: geolocation data (delayed some hours)	NTUA, OASA & LIT
	Weather data: hourly weather recordings and weather events	All partners
Madrid Pilot	Demographics data on Spain regions: yearly updates	All partners
	Mobile network operator data (anonymised)	Nommon
	All train and metro stations	All partners
	Shared mobility providers: number of trips, number of vehicles used per day, number of vehicles available in the city (pay-per-use API)	Nommon
	CRTM Madrid Regional Transport data: Line route, vehicles and timetables for the different public transport systems in Madrid (metro, cercanías, city and intercity buses, …)	All partners
	Regional transport survey data: user mobility preferences/patterns/behaviour	All partners
	AEMET weather: hourly weather records & weather events.	All partners
Almelo Pilot:	Road telematics/sensor data: inductive loops and traffic signal data, 5 years of historic data	UTWENTE
	Fleet Vehicle data: fleet operations data	UTWENTE
	CCTV: Real-time vehicle sensor/detection data	UTWENTE
	Radar data: real-time data on detecting bicycles	UTWENTE
	Weather KNMI data: hourly weather recordings	All partners

	Weather forecast data (Buienradar/Buienalarm): hourly weather forecasts for up to 14 days (commercial API)	All partners / Public
	Public transport data (GTFS): routes and schedules of buses	All partners
Slovenian Pilot:	GoOpti demand data since 2020 for the region of Slovenia + neighbouring connecting routes (all the close by airports and major cross border cities)	GoOpti, JSI
	Ljubljana public transport routes and schedules	All partners
	Weather data (weatherAPI.com): hourly weather records and weather events	All partners
	Flights information data: flights schedules data for all airports on the GoOpti routes: Ljubljana, Zagreb, Trieste, Venice, Vienna, Munich (++). Historic data available to all partners, real-time data available on commercial API (pay-per-request).	All partners
	OpenStreetMaps (OSM) data: data for routes construction, used in optimization algorithms.	
	Google maps data: data for routes construction, used in optimization algorithms. Openly available data on Google API	All partners
	DARS data: traffic data on backbone road infrastructure (highways): traffic events, road conditions, traffic counters.	All partners

During the project, additional traffic-related data will be generated covering different aspects (i.e., transport, environment, economy, society, etc.),

All results (except for administrative data containing legal, financial or sensitive information) will be published under a CC-BY 4.0 licence using the recommended file formats and metadata. The file formats used will be open and allow for straightforward reuse. Where possible, metadata will follow the widely accepted Dublin Core and DataCite specifications. The project will use reference metadata. In the ZENODO upload interface, all required fields will be filled in, including the 'upload type' (e.g., 'publication/project deliverable' for deliverables, 'publication/project milestone'' for milestones, 'dataset' for the underlying data), the full list of 'authors' with a listing of contributing authors in order of lead author (with ORCIDs if available), the 'description' and the 'version'. The following recommended/optional fields will also be filled in: 'Funding' (European Commission with value 101077049), 'Related/alternate identifiers', 'Contributors' and 'Subjects'.

In addition, the official document “Horizon Europe Data Management Plan Template” requests a set of specific questions that we are addressing below:

- *Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.*

In UC1 Athens pilot, we will re-use loop detector data and public transit supply data. We will re-use them to optimise the public transit service of Athens in terms of synchronisation of transfers between metro and bus transport modes. We will also use historical weather data to identify the spatio-temporal passenger demand variation due to changes to the weather conditions.

In UC1 Almelo pilot, existing data will be re-used for the activities of CONDUCTOR. Existing data that will be re-used are road telematics/sensors data from inductive loop detectors and traffic signals, and public transport data for the set-up, validation and calibration of the (on-the-side) simulation. The data sources will be combined with re-used historical weather data as to predict spatio-temporal variations in urban traffic conditions. Historical fleet operations data are used for testing purposes.

In UC2, we will use the following existing data: weather, historical demand data from GoOpti. It will be used to model demand prediction and mobility patterns. The initially planned use of public transportation data will probably not be needed for modelling.

In UC3, mobility indicators extracted by Nommon from the fusion of different big data sources will be re-used. This will be at two levels: (i) Within the CONDUCTOR project, these data will first be used to develop technical solutions and then to calibrate and feed the models that will be used in the UC; (ii) After or outside the CONDUCTOR project, these data will be stored in the Nommon database and used for further development of Nommon solutions. Regarding the data used by Aimsun for the development, calibration and validation of the models, two Aimsun-based models should be distinguished: (i) M-30 Aimsun microscopic model, which covers only the M-30 ring road network, is owned by Aimsun. This data will be re-used by Aimsun within the CONDUCTOR project, to simulate and integrate the proposed methods and solutions. Outside of the CONDUCTOR project, the data will be stored in Aimsun database to be used for future developments. (ii) Madrid Aimsun macroscopic model, which covers the entire urban area of Madrid at a macroscopic level, is owned by the Madrid City Council. Aimsun already has access to their model, but discussions are ongoing to obtain permissions to use it in CONDUCTOR. Aimsun will re-use this data for the CONDUCTOR project and Madrid City Council will re-use this data after CONDUCTOR project for future exploitation of the model.

- *What types and formats of data will the project generate or re-use?*

The formats of the most data generated or re-used are txt, json, csv, and ang.

In addition, in UC1 Athens pilot, the generated data will be the outputs of microscopic simulation, such as travel time and distance, delay time, density, flow, speed, queues, number of stops, etc., as well as emissions of CO2, NOx, and PMs.

For UC1 Almelo pilot, re-used data will mainly be V-LOG data. V-LOG is the open protocol used in the Netherlands for logging all signalized intersection-related events, such as signal requests, signal realizations and interruptions, inductive loop detections, etc.

- *What is the purpose of the data generation or re-use and its relation to the objectives of the project?*

In requirements gathering process, the main data sources were the survey respondents.

In UC1 Athens pilot, the generated data will be used to calculate the values of KPIs and for the validation and assessment of the Use Case in Athens. The re-used data will facilitate the development of vehicle scheduling and traffic management models.

In UC1 Almelo pilot, the re-used data will mainly be used for the set-up of a simulation environment, in which the developed and upgraded functionalities are tested before real-world tests occur. Data generation occurs for testing purposed and will be used in the real-world pilot to support decision making, thereby allowing a seamless journey for heavy duty vehicles along the arterial.

In UC2, demand predictions will be used to manage fleet operations: Creation of route plans for the next days.

In UC3, all the generated data will be used to feed the existing Aimsun Next simulation model. This model will be used to perform the traffic simulations for the analysis of the proposed solutions for the integration of the last mile delivery into the passenger transport services.

- *What is the expected size of the data that you intend to generate or re-use?*

In UC1 Athens, passenger demand, telematics (vehicle positioning), loop detector and weather data are planned to be used for the city of Athens. The expected size is 200MB-1GB per scenario.

In UC1 Almelo, the historical data for signalized intersections is expected to be 20 GB. Generated data is expected to be 10GB per analyzed scenario.

In UC2, demand prediction is planned for the region of Slovenia and northern Italy. The data for a daily demand represents a JSON object at the 1 MB level.

In UC3, the expected size is about 10 GB per analysed scenario.

- *What is the origin/provenance of the data, either generated or re-used?*

In requirements gathering process, the main sources were the survey respondents.

In UC1 Athens pilot, the data comes from a variety of data sources, including open data portals, private data from OASA and simulation.

UC1 Almelo uses a variety of data sources, including open data (e.g., weather and public transport data), road authority-owned data (e.g., from inductive loop detectors) and private data (fleet operations data). The generated data originates from University of Twente.

In UC2, the data is generated by the demand prediction model, accessed through the REST API service.

In UC3, the data (re-used or generated) comes from a variety of data sources including: open data portals, private data from Aimsun or Nommon from commercial agreements, private data from the municipality, internally generated data from Nommon or Aimsun.

- *To whom might your data be useful ('data utility'), outside your project?*

The data can be useful for anyone involved in modelling traffic events and processes. Mobility indicators and network data infrastructure, incident data, are useful for various stakeholders involved in traffic, transport and mobility planning, management, monitoring and provisioning. Some examples are: Madrid City Council or the Madrid Regional Transport Consortium, On-demand Mobility services in Athens, Traffic Management Centers in Athens, etc.

3 FAIR DATA

3.1 Making data findable, including provisions for metadata

All data will have an associated metadata document (stored as a .txt file) that describes the most important aspects of the data. Project deliverables will be assigned a unique identifier: **CONDUCTOR_[deliverable number]_[deliverable name]_[version]_[date of submission],**

e.g., CONDUCTOR_D7.1_Management_handbook_v1.0_20221231 (already submitted).

- *Will data be identified by a persistent identifier?*

The data stored in ZENODO will have the Digital Object Identifier (DOI) assigned.

- *Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.*

In order to analyse and classify all the information provided by the different data sources, a factsheet has been designed (part of activities in T1.4). This factsheet includes the following metadata: general information (data source name, contact information, last factsheet update), data provider, data ownership, data usage, details (data source description, size, availability, lifetime, etc.), and field description. At the initial stage of the project theses sheets are internal documents for traceability purposes and will be completed as the data sources become available. Also, a general CONDUCTOR data sheet was generated in which data are organised by UC and category. A factsheet for each data source used is completed and shared with all the members of the project consortium. The factsheet is completed by the first partner that identifies the data source. Where needed, we will provide Swagger documentation on how to use the services, as well as description on semantics of the data service provided.

- *Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?*

We will provide search keywords for optimizing possibilities for data re-use.

- *Will metadata be offered in such a way that it can be harvested and indexed?*

Metadata will be provided to enable harvesting and indexing. In addition, Swagger documentation will be available.

3.2 Making data accessible

CONDUCTOR commits to early and open sharing of all its results. Each partner will store and archive data in a trusted repository suggested by its institution. However, the project openly accessible outcomes will be deposited in ZENODO, where all meta data is openly available under CC0 licence, and all content is openly accessible through open APIs. ZENODO helps researchers to receive credit by making the research results citable and through OpenAIRE integrates them into existing reporting lines to the European Commission and other funding agencies. Citation information is also passed to DataCite and onto the scholarly aggregators. Zenodo provides DOI to all uploaded data, therefore the repository ensures that the data is assigned an identifier and resolves the identifier to a digital object.

All project outputs - documents and data of all types - will be open for comment, deposited in ZENODO, and be published on the project web platform as soon as they are available (with the

exception of individual questionnaires, which will be stored at each partner's premises). To ensure Open Access, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository. For scientific publications this happens at the time of publication at the latest. Anonymised survey data and software code (with supporting documentation) will be published in certified digital repositories and/or ZENODO. Source code will be shared via GitHub or alike. Data will be published using standard file formats (txt, pdf, csv etc.). All data will be accessible using standard tools.

The only data which will not be made openly accessible will be data which contains personally identifiable information (e.g., individual evaluation forms) and deliverables that are covered by confidentiality. The personal data processed in the project will not be made publicly accessible but kept closed and inaccessible to third parties.

3.2.1 Repository

- *Will the data be deposited in a trusted repository?*

During the duration of the project, data will be deposited in the ZENODO repository whenever possible, either in full or in the form of metadata, in order to improve findability. In the case of UC3, this only applies to the generated data that can be shared, as all shareable input data is already available in public repositories. Personal data will be stored on a local, secured server of the partner who is responsible for maintaining this data. Data from the requirements gathering process is stored internally in company repository (Google Drive).

- *Have you explored appropriate arrangements with the identified repository where your data will be deposited?*

The arrangements are in progress and will be presented to all partners in due course.

- *Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?*

Each upload to ZENODO is automatically assigned a DOI, to make item citable and traceable.

3.2.2 Data

- *Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.*

In general, data will be openly accessible whenever possible. At this stage of the project, we are still gathering information about the different data sources, so we do not know the privacy terms for some data sources. There are no restrictions on the use of openly accessible data, apart from giving appropriate credit to the originator of the data. In case the CONDUCTOR partners need to exchange any sensitive (personal, industrial) data, encrypted tools and secure FTP solutions provided by their IT services will be used.

The personalised data from the process of requirements gathering is only accessible internally to the consortium. Accessibility for others during the project is done upon request.

In UC1, public transit demand and telematics data will be sensitive. The data is provided for the scope of this project. They will be used to optimise the public transit service of Athens, Greece. The generated data from the project will, however, be publicly available.

In UC2, all data services and associated data will be openly accessible. For business reasons, no historical data will be exposed.

In UC3, the raw mobile network data and the individual activity and travel diaries calculated from it can only be used by Nommon on the basis of a private agreement with the mobile operator (Orange). The aggregated mobility indicators generated from these data can only be used by the project members under a consent agreement. The raw data of the shared mobility services will also not be publicly available. Access to them has yet to be determined with the data provider. However, as in the case of the mobile network data, the aggregated indicators extracted from it will be shared with the partners of CONDUCTOR. Aggregated mobility indicators will be made available for research purposes via an open data repository. The Aimsun M-30 microsimulation model will have restricted access. The Aimsun Madrid macroscopic model can only be accessed by Aimsun, due to private agreements with the Madrid City Council. The results/outputs of the simulation model - more precisely UC3 - will be shared with the CONDUCTOR partners. The Aimsun APIs developed during the project will also be shared with all partners.

- *If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*

No data embargo is foreseen by the CONDUCTOR consortium.

- *Will the data be accessible through a free and standardized access protocol?*

The data will be accessible via the ZENODO Repository services. Some locally stored data will be accessible via dedicated REST API & Swagger.

- *If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?*

Apart from the business and proprietary reasons described above (where data will not be accessible), no other restrictions are currently envisaged by the CONDUCTOR consortium.

- *How will the identity of the person accessing the data be ascertained?*

No restrictions on data access in repositories is envisaged by the CONDUCTOR consortium.

- *Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?*

Considering the type of data collected and generated by the CONDUCTOR project, there is no need for the establishment of a data access committee. Nevertheless, if the situation will require it, the consortium will consider setting up a data access committee composed of one member from each partner organisation in order to look at aspects such as user profiles, systematic monitoring of individuals, processing of large-scale data or intrusive methods of data collection (such as surveillance, geolocation tracking, etc.).

3.2.3 Metadata

- *Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?*

All metadata deposited in ZENODO will be openly available under the CC0 licence. They will contain enough information (including a direct link) to allow a user to access the data.

- *How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?*

Data will remain available and findable through the ZENODO repository, which guarantees the long-term stability of data identifiers and data availability. All metadata is guaranteed to remain available after data is no longer available.

- *Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?*

No special software is foreseen to access the available and open data.

3.3 Making data interoperable

- *What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?*

To ensure maximum compatibility and accessibility, the structured data from CONDUCTOR will be formatted in widely used formats, avoiding potential problems related to vendor lock-in. In addition, each dataset will include standard ZENODO metadata, making it easier to find and use the data. This approach not only facilitates data sharing and collaboration, but also ensures greater transparency and reproducibility in research.

In UC1, the data generated will be formatted based on a tabular structure. Therefore, the generated files will be encoded in CSV format, which is a lightweight format for exchanging this type of data, and we will specifically use the formal description defined in RFC 4180 standard.

Some of the UC2 data sources use proprietary schemas, though we will also include traffic infrastructure data in the DATEX II format as an interoperable standard.

The data generated by UC3 will be formatted so that it can be reused by third parties using automated processes. Most of these data are expected to have a tabular structure. Therefore, the generated files will be encoded in CSV format, which is a lightweight format for exchanging this type of data, and we will specifically use the formal description defined in RFC 4180 standard. In the case of visualisations and other material created for human understanding, special efforts will be made to provide the data presented in such a way that it can be easily ingested and processed by computer systems. Data files shall be accompanied by metadata describing each dataset and the columns it contains.

- *In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?*

It is not intended to use uncommon specific ontologies.

- *Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?*

In UC1, we will refer to the data provided by loop detectors.

Some of the data in UC2 will also include data from other/public traffic infrastructure APIs.

3.4 Increase data re-use

- *How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?*

Each dataset will be enriched with documentation to validate the data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.). The results of the descriptive analysis of the requirement gathering process performed with Jamovi (version 2.2.5) and the datasets (with legends for coding the variables) will be provided in pdf files. The Swagger documentation will include an additional description of the data semantics. In addition, the required information will be included in the public technical deliverables of WP3 and WP5.

- *Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?*

Most data will be openly exposed. Open data will be licenced under the Creative Commons licence CC BY-NC-SA 4.0, i.e., Attribution-NonCommercial-ShareAlike 4.0 International.

- *Will the data produced in the project be useable by third parties, in particular after the end of the project?*

Most data from CONDUCTOR will be made freely available (via ZENODO). For shared information, a standard format, open-source software and appropriate documentation will guarantee re-usability by third parties.

- *Will the provenance of the data be thoroughly documented using the appropriate standards?*

All data managed under the project will be licenced using standard re-use licences in accordance with the obligations set out in the Grant Agreement and will be treated in accordance with the requirements of EU GDPR.

- *Describe all relevant data quality assurance processes.*

A quality assurance process will be carried out during the duration of the project. The origin of the data will be carefully documented following the Dublin Core Model. All data produced will be assessed in a quality assurance process to avoid inconsistencies and other anomalies in the data. Data cleansing tasks will also be performed to improve data quality. The process may involve automated and manual procedures, including: (i) statistical analyses such as frequencies, means, ranges or clustering to detect errors and anomalous values; (ii) checking values that are out-of-range; (iii) checking completeness of data; (iv) adding variable and value labels where appropriate.

- *Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.*

In addition to the data, all other relevant research findings will also be taken into account and stored appropriately.

4 OTHER RESEARCH OUTPUTS

- *In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).*
- *Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.*

In CONDUCTOR, both digital and physical outputs will be generated. Their compliance with the principles of FAIR will be managed by their creators and documented in further versions of this DMP.

5 ALLOCATION OF RESOURCES

- *What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?*

The cost for making data FAIR is estimated to be zero. ZENODO is funded by the European OpenAIRE program, which means that archiving data in this repository will be free of charge. However, the repositories of certain institutions is not free of charge but the partners were asked to budget such costs when preparing the CONDUCTOR proposal.

- *How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)*

Each partner will bear the costs of making their data and other research outputs FAIR in CONDUCTOR. The costs of publishing results produced by several partners will be shared or covered by the whole consortium, if the results are of general interest. During the project, these costs will be covered from the project budget. How the costs are to be covered after the end of the project, needs to be discussed in the first period of the project.

- *Who will be responsible for data management in your project?*

Each beneficiary leading the work package is responsible for the preparation of the datasets to make the data collected under its own activities FAIR; provided that JSI, as Technical Lead, shares the responsibility for the overall coordination and monitoring. In addition, the consortium partners are responsible for ensuring that their activities comply with all applicable local, government and international laws, regulations and guidelines.

- *How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?*

Long-term preservation and needed resources (e.g., costs and potential value, who decides and how, what data will be kept and for how long) will be discussed during the CONDUCTOR project and will be described in the next version of the DMP.

6 DATA SECURITY

- *What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?*
- *Will the data be safely stored in trusted repositories for long term preservation and curation?*

Each partner is responsible for implementing all necessary measures to secure the processing of research data required for the implementation of its project activities. In particular:

- Taking technical and organisational measures to prevent unauthorised access to datasets and electronic files, e.g., through login restrictions and the use of strong passwords;
- Establishing clear access rules to their databases (e.g., based on predefined user lists and restrictions on access rights to prevent the possibility of copying, printing or downloading);
- Performing processing (also on behalf of a project partner) under the best possible control of the data and ensuring appropriate security measures;
- Setting local encrypted storage (e.g., enabling full hard disc encryption, which puts the data in a form that makes it unintelligible);
- Protecting data transmission (e.g., using HTTPS communication via Hypertext Transfer Protocol - HTTP within a connection encrypted by Secure Sockets Layer - SSL).

The above measures are in line with the legal provisions of the Consortium Agreement regarding specific responsibilities in relation to data protection (Article 15) and non-disclosure of information (Article 13). However, there are specific legal requirements for the processing of personal data that must be fulfilled by all project partners in accordance with the applicable EU and national legislation on the protection of personal data. It is important to consider that the processing of personal data is not only necessary during the implementation of the project but may also be necessary for post-project exploitation. Consequently, project partners must take all measures to ensure adequate and effective compliance with data protection regulations during and after the project. This applies to the notification procedures (i.e., Privacy Policy and Data Processing Agreements) and security measures.

Each partner is responsible for its own compliance with the legal requirements and provisions of the EU General Data Protection Regulation (GDPR). An assessment of the GDPR compliance measures will be carried out during the project in order to verify the adequate compliance with the GDPR during the project implementation. The assessment of the GDPR compliance measures will be carried out by JSI among the project partners processing personal data required for the project implementation.

7 ETHICS

- *Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).*
- *Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?*

Partnerships and co-creation are at the heart of CONDUCTOR. For the project to be successful, it is important to collect data about the participants and stakeholders and their ideas, interests and opinions. The partners collect this data using a range of techniques and in compliance with relevant data protection regulations.

CONDUCTOR reaches out to many organisations, individuals and other projects and organises interviews, workshops and other activities involving people inside and outside the project. In doing so, CONDUCTOR ensures proper handling of ethical aspects, values and data protection in accordance with Articles 14 and 15 of the Grant Agreement. While each partner is responsible for their own actions, INTRA as coordinator of the project CONDUCTOR guides and supports partners to act according to the DMP.

Personal data of the project partners used to manage the project activities will not be shared with third parties. The CONDUCTOR website is managed by JSI, the WP leader of Dissemination, Exploitation and Communication. The database of CONDUCTOR contacts is managed by INTRA. The protection of personal data is in line with the GDPR.

8 OTHER ISSUES

- *Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?*

CONDUCTOR partners agree to the principles and good practices of the European Code of Conduct for Research Integrity (ALLEA, 2017). In addition, project partners may have their own national and/or institutional data management policies for the data entrusted to them. All project partners are based in the EU.

In the next version of the DMP, the information on all national/state/sectoral/departmental procedures for data management will be made available.

9 CONCLUSIONS

The Data Management Plan will be modified and updated throughout the course of the project whenever significant changes occur, such as new data or changes in the consortium's policy and the progress of the dissemination and exploitation strategy for each dataset collected or generated.

At the end of the project, the final version of the DMP will be published as D7.3 *Data Management Plan - Final version*.

A. ABBREVIATIONS AND DEFINITIONS

Term	Definition
AEMET	**A**gencia **E**statal de **Met**eorología
ALLEA	**All E**uropean **A**cademies
API	**A**pplication **P**rogramming **I**nterface
CCAM	**C**ooperative, **C**onnected and **A**utomated **M**obility
CCTV	**C**losed-**C**ircuit **T**elevision
CSV	**C**omma **S**eparated **V**alues
DARS	**D**ružba za **A**vtoceste v **R**epubliki **S**loveniji
DATEX II	**Dat**a **Ex**change standard
DMP	**D**ata **M**anagement **P**lan
DOI	**D**igital **O**bject **I**dentifier
FAIR	**F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability
FTP	**F**ile **T**ransfer **P**rotocol
GA	**G**rant **A**greement
GDPR	**G**eneral **D**ata **P**rotection **R**egulation
GTFS	**G**eneral **T**ransit **F**eed **S**pecification
HTTP	**H**yper**t**ext **T**ransfer **P**rotocol
HTTPS	**H**yper**t**ext **T**ransfer **P**rotocol **S**ecure
JSON	**J**ava**S**cript **O**bject **N**otation
KNMI	**K**oninklijk **Ne**derlands **M**eteorologisch **I**nstituut
ORCID	**O**pen **R**esearcher and **C**ontributor **ID**
OSM	**O**pen **S**treet **M**ap
REST	**RE**presentational **S**tate **T**ransfer
SSL	**S**ecure **S**ockets **L**ayer
UC	**U**se **C**ase
WP	**W**ork **P**ackage