



HARMONY
SPATIAL & TRANSPORT PLANNING FOR A NEW MOBILITY ERA

Holistic Approach for Providing Spatial & Transport Planning Tools and Evidence to Metropolitan and Regional Authorities to Lead a Sustainable Transition to a New Mobility Era

D10.3 Data Management Plan Version 1

Submission date: 31/03/2020



[@Harmony_H2020](https://twitter.com/Harmony_H2020)

[#harmony-h2020](https://twitter.com/Harmony_H2020)



<https://www.linkedin.com/company/harmony-h2020/>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815269



SUMMARY SHEET

PROJECT

Project Acronym:	HARMONY
Project Full Title:	Holistic Approach for Providing Spatial & Transport Planning Tools and Evidence to Metropolitan and Regional Authorities to Lead a Sustainable Transition to a New Mobility Era
Grant Agreement No.	815269 (H2020 – LC-MG-1-2-2018)
Project Coordinator:	University College London (UCL)
Website	www.harmony-h2020.eu
Starting date	June 2019
Duration	42 months

DELIVERABLE

Deliverable No. - Title	D10.3 – Data Management Plan Version 1
Dissemination level:	Public
Deliverable type:	Report
Work Package No. & Title:	WP10 - Dissemination, Exploitation & Innovation Management
Deliverable Leader:	ICCS
Responsible Author(s):	Efthimios Bothos (ICCS), Babis Magoutas (ICCS), Gregoris Mentzas (ICCS)
Responsible Co-Author(s):	Ioannis Tsouros (UAEGEAN), Panos Georgakis (UoW), Angelica Salas (UoW), Francesca Fermi (TRT), Athina Tsimpa (MOBY), Ilias Gerostathopoulos (MOBY), Nellie Tzivelou (OASA), Michiel de Bok (TUD), Ioanna Kourounioti (TUD), Giuseppe Estivo (CDT), Andrea Marella (AUCM), Sridhar Raman (OCC), Economides George (OCC), Lampros Yfantis (UCL)
Peer Review:	Panagiotis Georgakis (UoW)
Quality Assurance Committee Review:	
Submission Date:	31/03/2020

DOCUMENT HISTORY

Version	Date	Released by	Nature of Change
0.1	15/01/2020	ICCS	Initial Version with ToC
0.2	14/02/2020	ICCS	Data templates added
0.3	02/03/2020	ICCS	Information about data security added

0.7	18/03/2020	ICCS	First consolidated version for the consortium.
0.8	23/03/2020	ICCS	Internal review comments addressed
1.0	31/03/2020	ICCS	Final version for submission

1 TABLE OF CONTENTS

EXECUTIVE SUMMARY	4
1 Introduction	5
2 Data Summary	6
2.1 Purpose of Data Collection	6
2.1.1 Data required to build the HARMONY Models	6
2.1.2 Data emerging from the Pilot Studies	6
2.2 Relation of Data to the project's objectives	6
2.3 Types and formats of the project's data	6
2.3.1 Primary data from surveys.....	7
2.3.2 Primary data emerging from HARMONY workshops	7
2.3.3 Secondary data from the pilot sites	7
2.3.4 Data generated by the models	14
3 Fair Data	15
3.1 Making data findable, including provisions for metadata	15
3.1.1 Naming Convention Strategy.....	15
3.1.2 Version Numbering Strategy	15
3.1.3 Metadata & Search keywords	16
3.2 Making data openly accessible	16
3.2.1 Datasets.....	16
3.2.2 Scientific Publications	16
3.2.3 Source code	16
3.3 Making data interoperable	17
4 Allocation of Resources	17
5 Data Security and Protection	17
5.1 Storage of sensitive data	17
5.2 Adherence to the General Data Protection Regulation	17
6 Ethical Aspects	19
6.1 Informed Consent	19
6.2 Exchanging, archiving and preservation of data	19
7 Conclusions	20
8 APPENDIX I: Secondary data definition template	21
9 APPENDIX II: Model Outputs data definition template	28

LIST OF ABBREVIATIONS

Abbreviation	Explanation
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Re-usable

EXECUTIVE SUMMARY

Management of data is an important element of large scale multi-disciplinary projects. As such, HARMONY will be collecting, using and generating a heterogeneous set of data throughout its lifecycle. This deliverable is the initial version of the project's Data Management Plan and explains the proposed actions for the overall control of HARMONY's data and publications. The deliverable embarks with a summarisation of the types and sources of data and continues with the description of the practices for safeguarding that the HARMONY research data are findable, accessible, interoperable and re-usable (FAIR). An account of the allocated resources for data management is included. The deliverable concludes with an appreciation of the data security and ethical considerations principles to be adopted.

1 Introduction

The Data Management Plan (DMP) explains the proposed actions for the overall control of HARMONY's data and publications. The DMP is a "living document" that will be constantly updated during the project. More specifically, two further versions (2 and 3) of the DMP will be produced in M30 and M42 respectively.

In order to derive the data sources which will be used and created within the HARMONY project a collaborative methodology is followed where all partners dealing with data are involved. Templates to record existing and new datasets were created and were provided to relevant partners. More specifically, one template was provided to pilot cities to gather available datasets and another was provided to transport modellers to gather the data that will be the outputs of the different models. The templates are provided in Appendix 1 and include information regarding the dataset's description, purpose and utility, reference and name, partner involved, format, related metadata and standards, relation to the project's objectives, whether it is a new or existing dataset and expected size. Partners were then asked to fill in the corresponding templates. The outcome of this process was a record of existing and new datasets which are described in Sections 2.3.1 and 2.3.2 of this deliverable. At this point datasets from the following partners were collected: UAEGEAN, UoW, TRT, MOBY, OASA, TUD, CDT, GROT, OCC.

The remainder of the deliverable is structured as follows. In Section 2 the document embarks with a summarisation of the types and sources of data and continues in Section 3 with the description of the practices for safeguarding that the HARMONY research data are findable, accessible, interoperable and re-usable (FAIR). An account of the allocated resources for data management is included. Then as an appreciation of the data security and ethical considerations, principles to be adopted is provided in Sections 4 and 5. The conclusions and next steps are part of Section 6.

2 Data Summary

2.1 Purpose of Data Collection

The goal of HARMONY is to develop a new generation of harmonised spatial and multimodal transport planning tools which comprehensively model the dynamics of the changing transport sector and spatial organisation, enabling metropolitan area authorities to lead the transition to a low carbon new mobility era in a sustainable manner. The HARMONY model suite is designed to assess the multidimensional impacts of the new mobility concepts and technologies. The model suite integrates: 1. land-use models (strategic/long-term), 2. people and freight activity based models (tactical/mid-term), and 3. multimodal network (operational/short-term) models allowing for vertical planning. This integrated approach is necessary for authorities to understand if policies are sustainable, while also contribute to meeting COP22 targets, social equality and wellbeing.

To achieve its goal, HARMONY makes use of existing data from cities where it is piloted, and generates new data in terms of travel surveys and actual deployment and use of new mobility services. These data are fed into the HARMONY models which in turn generate data for transport and spatial designs.

2.1.1 Data required to build the HARMONY Models

These data are used with the aim to build transport and spatial models at three levels: strategic, tactical and operational and include both primary data, i.e. data collected in the course of the project and secondary data, i.e. data that have been collected for some other purpose outside the project's scope, but can be utilized for feeding the HARMONY models. More specifically the following data types have been identified:

- Primary data emerging from the HARMONY surveys
- Primary data emerging from HARMONY workshops in the form of requirements
- Secondary data from the HARMONY pilot sites
- Data generated by the models (outputs) and are provided as input to the different levels or enable transport modellers to extract KPIs of interest

2.1.2 Data emerging from the Pilot Studies

HARMONY will apply the HARMONY model suite in six (6) pilot studies (Athens, Trikala, Katowice, Oxfordshire, Turin, Rotterdam). It is expected that the models will generate evidence which will allow the identification of sustainable and eco-rational mobility solutions, measures, policies, and business models to address current and future challenges in metropolitan areas.

2.2 Relation of Data to the project's objectives

The following table summarizes the relation of the different data categories to the project objectives. Note that O1, O2, O3, O4 are related to the implementation of the HARMONY models whereas O5, O6, O7, O8, O9 are related to the application of the models and dissemination of results.

O1, O2, O3, O4	Data required to build the HARMONY models, including: <ul style="list-style-type: none"> - Primary data emerging from the HARMONY surveys - Primary data emerging from HARMONY workshops in the form of requirements - Secondary data from the HARMONY pilot sites - Data generated by the models and are provided as input to the different levels
O5, O6, O7, O8, O9	Data emerging from the HARMONY pilot studies

2.3 Types and formats of the project's data

In the following sections, we describe a number of datasets which have currently been identified within the HARMONY project. Note that the datasets will be updated as the project advances and further information will be added to this document.

2.3.1 Primary data from surveys

A number of surveys are planned within the HARMONY project to gather data from travellers which will support the development of the models. These include:

- Passenger survey in Oxfordshire and Turin: for these surveys the software-as-a-service smartphone-based travel survey tool of MOBYx will be used to host the HARMONY questionnaires and collect the travel data. In order for the participants to use the MOBYapp smartphone app, they have to create an account, where they will be asked to provide their e-mails. The MOBYapp tracks the travel patterns of the individuals. In addition, a further validation of the tracking data is required, where the individuals will be asked to verify the type of the location (i.e. home, work, shopping etc.).
- Passenger survey in Trikala: in order to materialise the demonstration with drones delivering medicines from the city to the rural areas, we need to have the name, the home address and the home phone number of the individuals that will participate in the demonstration (address is needed to navigate the drones). Due to the fact that most of these participants are elderly and technology-illiterate, only personal interviews with paper-based questionnaires will be conducted.

2.3.2 Primary data emerging from HARMONY workshops

The project will organise a series of stakeholder engagement events (WP1, WP9, WP10) and primary data collection from passengers and freight actors (WP9). The data emerging from the workshops provide requirements for the implementation of the HARMONY models.

2.3.3 Secondary data from the pilot sites

Secondary data are provided by the pilot partners and will be used as input to the HARMONY models. At this point the consortium is in the process of collecting the information for the available secondary data including information for accessing the data and related access rights. The next step involves an evaluation regarding their fit-for-purpose for the HARMONY models. The following tables provide an overview of the available datasets listed by the pilot partners as of M10 of the project. Note that this list is continuously updated and revised. An updated list will be provided in the next version of D10.3.

2.3.3.1 City of Turin, Italy

2.3.3.1.1 Land Use Data

Data Type	Data Format	Data Year	Description
Economic			
Employment	csv, shapefile	2011	National Census with local area disaggregation. Available at: http://dati-censimentoindustriaeservizi.istat.it/FileView2.aspx?IDFile=5355745c-6bfa-4663-b602-4d1c53991d46
Retail Activities	csv, shapefile	2011, 2016	National Census and Municipality Retail Office Available at: http://dati-censimentoindustriaeservizi.istat.it/FileView2.aspx?IDFile=ef2540e8-a694-4006-8a45-1219f9e05dc0
Land use			
Land Use Data	csv, shapefile	2019	Available at: https://webgis.arpa.piemonte.it/ags101free/rest/services/suolo/Consumo_Suolo_2017/MapServer/WMTS/1.0.0/WMTSCapabilities.xml
Demographics			
Population	csv, shapefile	2019	Census Regione Piemonte, Available at: https://www.regione.piemonte.it/web/sites/default/files/media/documenti/2019-09/elenco_comuni_agg.2019.xls
Population Projections	csv, shapefile	2019	Only global percentage rate (without any zonal disaggregation) ISTAT. Available at: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_PREVDEM1#
Occupational Class	csv	2017	ISTAT, Available at: http://dati.istat.it/
Ethnicity	csv	2019	ISTAT, Available at: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPSTRCIT1#
Income	csv	2017	Income groups for classes and types. Available at: http://dati.istat.it/
Urban density	csv, shapefile	2017	Income groups for classes and types. Available at: http://dati.istat.it/
Housing			
Housing Tenure	csv	2011	National Census with local area disaggregation. Available at: http://www.datigeo-piem-

			download.it/direct/Geoportale/RegionePiemonte/OMI/ValoriOMI2016_su_Edifici_oBDTRE2017.zip
House Prices	csv	2011	National Census with local area disaggregation. Available at: http://www.datigeo-piem-download.it/direct/Geoportale/RegionePiemonte/OMI/ValoriOMI2016_su_Edifici_oBDTRE2017.zip
Topography			
General topography	Shapefile	2019	Available at: http://www.geoportale.piemonte.it/cms/bdtre/modalita-di-pubblicazione-e-fruizione
Point of Interest Data	csv	2017	Available at: http://www.geoportale.piemonte.it/cms/bdtre/modalita-di-pubblicazione-e-fruizione
Digital Elevation Model	shapefile	2013	Maximum detail GSD: 1 m. Available at: http://wms.pcn.minambiente.it/ogc?map=/ms_ogc/WMS_v1.3/servizi-LIDAR/LIDAR_PIEMONTE.map
Administrative Boundaries			
Municipal Boundaries	Shapefile	2019	Available at: http://www.istat.it/storage/cartografia/confini_amministrativi/archivio-confini/non_generalizzati/Limiti_2016_ED50.zip
Local Authority Boundaries	Shapefile	2019	Available at: http://www.istat.it/storage/cartografia/confini_amministrativi/archivio-confini/non_generalizzati/Limiti_2016_ED50.zip
Building Data			
Building Footprints	Shapefile	2019	Building footprint outlines. Available at: http://webgis.arpa.piemonte.it/ags101free/rest/services/suolo/Consumo_Suolo_2017/MapServer/WMTS/1.0.0/WMTSCapabilities.xml
Building Floorspace / Heights / Storeys	Shapefile	2019	Estimated volumetric unit per building. Available at: http://www.geoportale.piemonte.it/cms/bdtre/modalita-di-pubblicazione-e-fruizione
Building Function	Shapefile	2019	Available at: http://www.geoportale.piemonte.it/cms/bdtre/modalita-di-pubblicazione-e-fruizione
State / Public Housing Estates	Shapefile	2017	Location of large public housing estates. Available at: http://geovector3.territorio.csi.it/OGCProxyV2/bdtre_imm?service=WFS&request=getCapabilities
Planning Policy Data			
Locations for major future urban development	Shapefile	1995	Only old data, currently under development
Planning zoning	Shapefile	1995	Only old data, currently under development
Environmental or other development restrictions	Shapefile	2017	-
Major future transport infrastructure development	Shapefile	1995	Only old data, currently under development. Available at: http://www.geoportale.piemonte.it/cms/bdtre/modalita-di-pubblicazione-e-fruizione
Other data			
Utility Networks	Shapefile	2019	Only river. Available at: http://www.geoportale.piemonte.it/cms/bdtre/modalita-di-pubblicazione-e-fruizione

2.3.3.1.2 Transport Supply Data

Data Type	Data Format	Data Year	Description
Network data			
GIS shapefile of TAZ system and georeferences of centroids and connectors	shapefile	2019	Available at: http://www.istat.it/storage/cartografia/basi_territoriali/WGS_84_UTM/2011/R01_11_WGS84.zip
GIS shapefile of study area	shapefile	2019	Available at: http://www.istat.it/storage/cartografia/basi_territoriali/WGS_84_UTM/2011/R01_11_WGS84.zip
Open Street Map data	shapefile, OSM file	2020	Available at: http://download.geofabrik.de/europe/italy/nord-ovest-latest-free.shp.zip
Aerial photography and/or CAD for refinements (Google Streetview can be a replacement for this requirement)	raster image	2015	2018 Municipality of Turin, 2016 Piedmont Region. Available at: http://www.geoportale.piemonte.it/geocatalogorp/index.jsp?geonetworkrp/srv/ita/metadata.show&uuiid=r_piemon:94379297-e72a-41f8-918d-f497a956eb39&currTab=rndt
Bike lane network	shapefile	2019	Available at:

			http://geovector3.territorio.csi.it/OGCProxyV2/bdt_re_viab?service=WFS&request=getCapabilities
Walking network	shapefile	2019	Restricted access
Intersection (node coordinates)	shapefile, Visum network	2019	Source: 5T, Restricted access
Number of lanes	shapefile, Visum network	2019	Source: 5T, Restricted access
Lane connection information	shapefile, Visum network	2019	Source: 5T, Restricted access
Link free-flow speed	shapefile, Visum network	2019	Source: 5T, Restricted access
Lane widths	shapefile, Visum network	2019	Source: 5T, Restricted access
Location of multimodal transshipment terminals	shapefile	2019	source: 5T, Restricted access
Airport location data	shapefile	2019	Restricted access
Traffic Control Data			
Sign data (e.g., location of sign data, stop, yield, exit signing and/or lane turning assignment)	shapefile, Visum network	2019	source: 5T, Restricted access
Fixed Signal Control data	shapefile, Visum network	2019	source: 5T, Restricted access
Signalized Intersections (for each signal groups, signal phases, control plans, coordinates)	shapefile, Visum network	2019	source: 5T, Restricted access
(Semi) Actuated Signal Control data	csv	2020	Only for Municipality of Turin, source 5T, Restricted access.
Loop detector data	csv, shapefile	2019	source: 5T, Restricted access.
Location of variable message signs and set of possible pre-fixed messages	csv, shapefile	2019	source: 5T, Restricted access.
Location of variable speed signs, operational rules and algorithms used	csv, shapefile	2019	source: 5T, Restricted access.
Public Transport Data			
GTFIS files availability	GTFIS format	2019	source: 5T, Restricted access.
Line routing (ideally GIS based)	shapefile	2019	source: GTT, Restricted access.
Stops location (ideally GIS based)	shapefile	2019	source: GTT, Restricted access.
Type of vehicles used to operate each line	csv	2020	source: GTT, Restricted access.
Parking Data			
GIS files availability	shapefile	2019	Only for Municipality of Turin, source 5T, Restricted access.
Parking space availability system	shapefile, csv	2019	Only for Municipality of Turin, source 5T, Restricted access.
Freight-related data			
Terminal locations	shapefile, csv	2019	-
Energy, emission, noise data			
Vehicle Engine Type data	csv	2018	Disaggregated for Municipalities; ACI, Available at: http://www.aci.it/fileadmin/documenti/studi_e_ricerche/dati_statistiche/Prov_C_ateg_2014.xls
Noise emission data	shapefile, csv		Maps, data of 3 monitoring stations, Available at: https://webgis.arpa.piemonte.it/Geoviewer2D/?config=other-configs/rumore_config.json
Classification of vehicles	csv	2018	Category 1: Light motor vehicles, Category 2: Medium heavy vehicles, Category 3: Heavy vehicles, Category 4: Powered two wheelers. Disaggregated for Municipalities; ACI Available at: http://www.aci.it/fileadmin/documenti/studi_e_ricerche/dati_statistiche/Prov_C_ateg_2014.xls

2.3.3.1.3 Transport Demand Data

Data Type	Data Format and source	Data Year	Additional Comments
Passenger Demand Data			
Household travel demand surveys	csv	2013	IMQ2013 (source AMP), Available at:

			http://mtm.torino.it/it/dati-statistiche/indagine-img-2013/base-dati-img-2013/IMG2013_opendata.zip
<i>Static OD matrix</i>	csv	2011	National Italian Census disaggregated for local area, passenger trips, period: from 7.00 to 8.00. Matrix from 5T by vehicle type, Available at: http://www.istat.it/storage/cartografia/matrici_pendolarismo/matrici_pendolarismo_2011.zip
<i>Static OD matrix per vehicle type</i>	csv	2019	source 5T: private car, freight vehicles (and public transport), Restricted access
<i>Data used for static OD matrix estimation and calibration</i>			1700 loop detectors, Restricted access.
<i>Skim matrices</i>	csv	2019	source 5T, travel time, distance and free flow, Restricted access
Pedestrian Data			
<i>Pedestrian counts at crossings per direction (estimated or observed)</i>	csv	2019	Only for main historical and commercial streets, Restricted access
<i>Bicycle counts or occupancy at exclusive bicycle lanes (in case of sharing lanes with other modes provide occupancy share)</i>	csv	2020	Max. 5 counts for entire municipality area, Restricted access
Other Data			
<i>Mobile phone data</i>	csv	2018	Provider: TIM, Restricted access

2.3.3.1.4 Calibration Data

Data Type	Data Format and source	Data Year	Additional Comments
Calibration Data			
<i>Loop detector data</i>	csv	2019	Traffic flow by vehicle type (car and trucks) and speed
<i>Loop detector data</i>	csv	2019	aggregation interval: 5 min
<i>Loop detector data</i>	csv	2019	time period: 2019, every day
<i>Loop detector data</i>	csv	2019	data status: processed data
<i>Loop detector location data</i>			number of loop detectors: 1700
<i>Travel time for routes with details of the measurement sampling and a clear description (ideally GIS based) of the routes</i>	csv	2018	DVI-TIM City Forecast
<i>Saturation flow data</i>	csv, shapefile	2019	From currently VISUM network
<i>Delay and queue data</i>			

2.3.3.1.5 New Mobility Services Datasets

Data Type	Data Format and source	Data Year	Additional Comments
Traditional Taxis (Cabs) - Supply data			
Fleet size	csv	2019	Restricted Access

Station-based or free-floating Carsharing (DriveNow, car2go, etc.) - Supply data			
Fleet sizes	csv	2020	Restricted Access
Fleet composition/Vehicle Types/Number for each type	csv	2020	Restricted Access
Vehicle Capacities	csv	2020	Restricted Access
Station capacity/dock-parking numbers	csv	2020	Restricted Access
Fuel Consumption	csv	2020	Restricted Access
Station-based or free-floating Bikesharing (Santander bikes, Lime, Ofo, OBike, etc.) -Supply data			
Fleet sizes	csv	2019	Restricted Access
Fleet composition/Vehicle Types/Number for each type	csv	2019	Restricted Access
Station capacity/dock-parking numbers	csv	2019	Restricted Access
Station-based or free-floating Scooters (Lime, VOI, etc.) - Supply data			
Fleet sizes	csv	2020	Restricted Access
Fleet composition/Vehicle Types/Number for each type	csv	2020	Restricted Access
Vehicle Capacities	csv	2020	Restricted Access
Station capacity/dock-parking numbers	csv	2020	Restricted Access
Fuel Consumption	csv	2020	Restricted Access
Station-based or free-floating Bikesharing (Santander bikes, Lime, Ofo, OBike, etc.)			
Static or dynamic daily station/zone stock level data	csv	2020	Restricted Access
Trips /ODs	csv	2020	Restricted Access
Rentals/Bookings	csv	2020	Restricted Access
Station-based or free-floating Scooters (Lime, VOI, etc.)			
Static or dynamic daily station/zone stock level data	csv	2020	Restricted Access
Trips /ODs	csv	2020	Restricted Access
Rentals/Bookings	csv	2020	Restricted Access

2.3.3.2 Oxfordshire County, UK

2.3.3.2.1 Transport Supply Data

Data Type	Data Format / Description
Network data	
GIS shapefile of TAZ system and georeferences of centroids and connectors	Shapefile
GIS shapefile of study area	Shapefile
Open Street Map data	Shapefile
Bike lane network	
Walking network	
Intersection (node) coordinates	Shapefile
Section: Road category, speed limit	Shapefile
Number of lanes	Shapefile
Lane widths	Shapefile
Airport location data	Shapefile
Traffic Control Data	
Fixed Signal Control data	Fixed control plans, traffic lights position, phases and groups; traffic control plan
(Semi) Actuated Signal Control data	Data to correctly code controllers (for adaptive signals): SCOOT, MOVA, etc.. + turn movement counts + the details of each phase must be obtained, such as minimum green time, maximum green time, yellow change interval time, and red clearance interval time.

Loop detector data	SCOOT loop locations
Ramp Meter Control Data	The metering rate (or headway) associated with a ramp meter and how the metering rate is determined (fixed, ALINEA, HERO, etc.)
Location of variable message signs and set of possible pre-fixed messages	Only one site in Oxfordshire which is owned and managed by HE but we do have data
Location of variable speed signs, operational rules and algorithms used	From Argonaut the UTM common database
	Road Safety team
Public Transport Data	
GTFS files availability	Compressed TEXT files updated as of 2018
Line routing (ideally GIS based)	Compressed TEXT files updated as of 2018
Stops location (ideally GIS based)	Compressed TEXT files updated as of 2018
Stops assigned to each line	Compressed TEXT files updated as of 2018
Timetable covering the simulation period/day	Compressed TEXT files updated as of 2018
Energy, emission, noise data	
Classification of vehicles	Category 1: Light motor vehicles, Category 2: Medium heavy vehicles, Category 3: Heavy vehicles, Category 4: Powered two wheelers ATC, Vivacity labs.

2.3.3.2.2 Transport Demand Data

Data Type	Data Format and source	Data Year	Additional Comments
Passenger Demand Data			
Household travel demand surveys			https://www.gov.uk/government/statistical-data-sets/ad-hoc-national-travel-survey-analysis
Static OD matrix	CSV files	JAN-FEB 2016	OD matrix data between wards of Oxfordshire. Sourced as part of Google's Better Cities project.
Pedestrian Data			
Pedestrian counts at crossings per direction (estimated or observed)	CSV files	AUG 2019-present	Data produced by Vivacity Lab camera sensors
Bicycle counts or occupancy at exclusive bicycle lanes (in case of sharing lanes with other modes provide occupancy share)	CSV files	AUG 2019-present	Data produced by Vivacity Lab camera sensors
Other Data			
Mobile phone data	CSV files		Data from INRIX

2.3.3.2.3 Data for New Mobility Services

Data Type	Data Format	Data Year	Additional Comments
Station-based or free-floating Bikesharing (Santander bikes, Lime, Ofo, OBIke, etc.) -Supply data			
Fleet sizes	JSON files	Few months in 2018	All dockless bike-sharing operators have ceased operations. There is some supply data from 2018.
Station capacity/dock-parking numbers	JSON files	Few months in 2018	

2.3.3.3 City of Athens, Greece

2.3.3.3.1 Land Use Data

Data Types	Data Format	Data Year	Additional Comments
Land use			

Land Use Data census 2011 land use classification of buildings

Demographics

Population census 2011

2.3.3.3.2 Transport Supply Data

Data Type	Data Format	Data Year	Additional Comments
Network data			
GIS shapefile of TAZ system and georeferences of centroids and connectors	Shapefile, MS-ACCESS file	2014	1284 internal zones plus 12 external zones.
GIS shapefile of study area	Shapefile	2014	Core rode network with directions and all necessary link data (speed, capacity and so on), coded turn prohibitions, coded traffic lights in some intersections (no signaling program coded)
Aerial photography and/or CAD for refinements (Google Streetview can be a replacement for this requirement)	JPG	2020	Obtained from VISUM aerial photography applet
Intersection (node) coordinates	Shapefile, CSV	2014	
Section: Road category, speed limit	Shapefile, MS-ACCESS file	2014	
Number of lanes	Shapefile, MS-ACCESS file	2014	
Link free-flow speed	Shapefile, MS-ACCESS file	2014	
Lane widths	Shapefile, MS-ACCESS file	2014	
Airport location daata	Shapefile, MS-ACCESS file	2014	
Traffic Control Data			
Fixed Signal Control data	Shapefile, MS-ACCESS file	2014	Traffic light positions, phases and groups, no signalling program
Signalized Intersections (for each signal groups, signal phases, control plans, coordinates)	Shapefile, MS-ACCESS file	2014	Traffic light positions, phases and groups, no signalling program
Public Transport Data			
GTFIS files availability	Shapefile, MS-ACCESS file	2020	data available for bus, trolleybus
Line routing (ideally GIS based)	Shapefile, MS-ACCESS file	2020	
Stops location (ideally GIS based)	Shapefile, MS-ACCESS file	2020	
Stops assigned to each line	Shapefile, MS-ACCESS file	2020	
Timetable covering the simulation period/day	Shapefile, MS-ACCESS file	2020	
Type of vehicles used to operate each line (provide fleet size and composition including length, number of doors and which are used for boarding, number of seats, passenger capacity)		2020	

2.3.3.3.3 Transport Demand Data

Data Type	Data Format and source?	Data Year	Additional Comments
Passenger Demand Data			
Household travel demand surveys	text file, csv	2006	household survey (socioeconomic data, trips, mode choice, trip purpose etc)
SP experiments (mode, route or vehicle purchase choice)	text file, csv	2006	stated-preference (time/cost) about mode choice
Static OD matrix	ms access	2014	24 h and peak period matrices for Private and Public Transport (both per trip purpose and aggregated)
Static OD matrix per vehicle type	ms access	2014	Private Vehicles, Public Transport
Skim matrices	ms access	2009	Time, distance, generalized cost
Freight Demand Data			
Freight OD matrix	ms access	2006	OD Matrix from 2006 survey

2.3.3.4 City of Rotterdam, Netherlands

2.3.3.4.1 Transport Supply Data

Data Type	Data Format	Data Year	Additional Comments
Network data			
GIS shapefile of TAZ system and georeferences of centroids and connectors	Shapefile		A zoning system was created for the study area consisting of 2049 zones The zoning is based on a publicly available administrative zone level ("buurten")
GIS shapefile of study area	Shapefile		Province of South-Holland
Freight-related data			
City constraints	shapefile		Regulations imposed by local governments (e.g. forbidding trucks on specific times, vehicle weight restrictions)
Locations for trans-shipment	spreadsheet		Transfer locations (intermodal facilities). This data is not available from a supply model but from a different dataset that was provided
Terminal locations	spreadsheet		
Customer locations	synthetic firm population		Synthetic population was created by TUD

2.3.3.4.2 Transport Demand Data

Data Type	Data Format and source	Data Year	Additional Comments
Freight Demand Data			
Truck trip diaries	Basibestanden goederenvervoer: SAV Automated XML trip data collection: ASCI		"Basisbestanden Goederenvervoer" and XML transport microdata are both collected by CBS, and consists of large numbers of truck trip diaries for the Netherlands. Source (micro)data are proprietary and only available for analysis under strict conditions
Freight OD matrix	ASCI		Freight vehicle OD matrix can be derived from regional transport model (1 truck type, 3 time periods) or from National Freight model (7 vehicle types, 10 NSTR, 24 hour). Freight vehicle OD matrix from regional transport model or national freight model can be made available by RWS
Freight OD matrix per mode and/or vehicle type Data used for static OD matrix estimation and calibration	ASCI		see above Road counts for freight vehicles. Count data are available from the regional transport model from RWS
Skim matrices	ASCI, Cube		Skim matrix can be derived from regional transport model (1 truck type, 3 time periods)

2.3.3.4.3 New Mobility Services Data

Data Type	Data Format and source	Data Year	Additional Comments
Traditional freight operators (Trucks)			
Fleet composition/Vehicle Types/Number for each type			aggregate descriptives
Zone to zone travel times and loading/unloading times			see transp_demand section

2.3.4 Data generated by the models

In this section a preliminary and limited list of datasets generated by the HARMONY models is provided. Note that the list is being updated and refined as the technical work progresses. A complete list will be provided in the next version of D10.3.

2.3.4.1 Employment Data

Description: The Regional Economy model generates future employment (including services, health and educational activities) which structure the demand for physical travel (used as input in the following models).

Partner: Provided by TRT.

2.3.4.2 Location quotient

Description: The Regional Economy model estimates location quotient by sector over time (used as input in the following models).

Partner: Provided by TRT.

2.3.4.3 Trip chains

Description: The multimodal freight controller will emulate the decision making of logistic operators, as well as behavioral factors that influence the choice and operation of innovative freight distribution approaches. It has as input the freight demand and it will generate the trip chains based on the integration of crowdshipping, microfreight logistics (e-cargo bikes, delivery bots, electric vans) and air transport (drones).

Partner: Provided by UoW.

2.3.4.4 Routes

Description: The freight controller receives real-time information from the simulator environment on congestion and delays to optimize freight operations. This data is used to generate/reschedule routes for the trips.

Partner: Provided by UoW.

3 Fair Data

3.1 Making data findable, including provisions for metadata

In order to make the project data findable, a Digital Object Identifier (DOI) will be requested for each artefact. In more details, DOIs from Crossref will be used for research publications, while DOIs from DataCite will be pursued for labelling each dataset of the project. In addition, a metadata record for each output of the project will be created and stored in the data directory. Amongst other fields, each metadata record will have a set of keywords that will make searches easier for external parties.

3.1.1 Naming Convention Strategy

In HARMONY, each data source will be provided with a specific name that is composed by different parts/elements, containing information about pilot country, data type or format and naming structure as follows:

ORIGIN_ORG_TOD_FORMAT_Info_VERSION

- **ORIGIN:** A prefix denoting if the dataset is pre-existing or new, followed by the first letters (three max) of the pilot's country (IT, UK, GR and NL) or GEN if the data artefact is pilot agnostic
- **TOD:** The type of data
- **FORMAT:** The data format/extension
- **Info:** Additional (abbreviated) information about the dataset. For example, the year when the dataset was published.
- **VERSION:** The version of the dataset.

3.1.2 Version Numbering Strategy

In HARMONY, we are following a data versioning strategy similar to software versioning, applying a two-part numbering rule: Major.Minor (e.g. V2.1). Major data revision indicates a change in the formation and/or content of a dataset that may bring changes in scope, context or intended use. For example, a major revision may increase or decrease the statistical power of a collection, require change of data access interfaces, or enable or disable answering of more or less research questions. A Major revision may incorporate:

- substantial new data items added to /deleted from a collection

- data values changed because temporal and/or spatial baseline changes
- additional data attributes introduced
- changes in a data generation model
- format of data items changed
- major changes in upstream datasets.

Minor revisions often involve quality improvement over existing data items. These changes may not affect the scope or intended use of initial collection. A Minor revision may include:

- renaming of data attribute
- correction of errors in existing data
- re-running a data generation model with adjustment of some parameters
- minor changes in upstream datasets.

3.1.3 Metadata & Search keywords

All datasets that will be openly available will be accompanied with metadata information which will render them findable by interested third parties. Search keywords will be defined and will be part of the related metadata for each dataset.

At this point we plan to use the CERIF <https://www.eurocris.org/cerif/main-features-cerif> metadata format. However, in the course of the project we will check and identify any other applicable formats.

3.2 Making data openly accessible

A number of datasets that will be used as part of the project will be offered by previous studies. Some of these datasets are already open to the public, while others are proprietary and have high commercial sensitivity. In the cases where private data are processed and aggregated (e.g. as part of a model, or functionality of a component) permission will be requested by the provider prior to making the altered data publicly available.

In reference to the nature of the user data involved, some of the results that will be generated by each project phase will be restricted to authorised users, while other results will be publicly available. As per our Ethics commitment during the negotiation phase of the project, data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU.

Since the DMP is expected to mature during the project, the subsequent releases of the deliverable will specify the repositories where the data will be stored and go into more detail on how this data can be accessed by the wider research community.

3.2.1 Datasets

Datasets characterised as “openly accessible” will be published in the following open repositories in OpenAire: <https://www.openaire.eu>

3.2.2 Scientific Publications

As required by the Grant Agreement, research publications will be made available through Green Open Access, where each publication needs to be made available at the HARMONY and Institutional portals. If applicable, Gold Open Access may be necessary, where the publication will be openly available through the publisher’s website. The publications of the project will be disseminated through the project’s dissemination and exploitation channels and follow the process described in the relevant project strategies.

3.2.3 Source code

It will be at the discretion of individual consortium members to decide whether the source code of their developed software is openly accessible. In such cases, different free and open-source software

licenses will be investigated and the appropriate ones will be selected. Open source code from the HARMONY project will be made available through a common GitHub Repository.

3.3 Making data interoperable

HARMONY partners will use metadata vocabularies when possible to render the provided datasets interoperable. The formats that will be used will be described in later versions of the data management plan.

4 Allocation of Resources

Regarding the resources related to data management activities, the project includes a work package that is responsible for data warehousing and travel surveys (WP3), a task for identifying pilot areas' data (T9.1) and a task for the data management plan. In total, ~105 person-months of effort have been allocated to the relevant WPs and Tasks. In addition, around €13,000 have been allocated for the generation of Open Access publications. The data management plan task is led by ICCS who together with all the partners will handle the management of data related to the technological aspects of the platform.

5 Data Security and Protection

The HARMONY platform will provide all required measures for secure data access with the usage of the latest encryption tools and protocols as well as data access control practices to prevent data misuse or manipulation. The data security mechanisms will be defined and implemented as part of WP3. It is envisaged that the starting candidates will be TLSv3 protocol for secure data connections and OAuth for access control.

5.1 Storage of sensitive data

Data privacy and user data protection issues will strictly follow the "user decides" principle. End-users will always have the possibility (and only the user) to decide which personal or private data to be used and all user referenced data will always be grouped and combined via anonymization tools to avoid the possibility of breaking it down to one user. All personal data stored within the HARMONY project will be archived for the lifetime of the project only, and will be coded, stored and kept privately in a secure location. No information will be shared with any external to the HARMONY consortium party without the prior express permission of the user. Sensitive information will be stored in an encrypted form, and all data will be protected by password access.

5.2 Adherence to the General Data Protection Regulation

The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679)¹ concerns issues related to the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). The regulation has been proposed and established by which the European Parliament, the Council of the European Union and the European Commission. It intends to strengthen and unify data protection for all individuals within the European Union (EU) and addresses issue related to the export of personal data outside the EU.

The GDPR aims primarily to give control to citizens and residents over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU. GDPR has been adopted on 27 April 2016, while it became enforceable from 25 May 2018, allowing a two-year transition period for member states. It is important to note that GDPR does not require national governments to pass any enabling legislation, and is thus directly binding and applicable. The HARMONY consortium is taking measures so that any user and personal data gathered from the project

¹ <http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679>

as well as related process adhere to GDPR. More specifically we have focused on twelve main steps which have been proposed by the ICO organization (Information Commissioner's Office) in the UK².

Step 1: Awareness. All partner organizations, corresponding decision makers and key persons within the HARMONY consortium have been informed of the GDPR enforcement and have been provided with related material in order to understand the impact of GDPR in their work. Partners will be requested to identify areas that could cause compliance problems under the GDPR and proceed to resolution actions if needed.

Step 2: Information held. The consortium, starting from this deliverable, is documenting the personal data that will be held along with information related to where these data came from and with whom they will be shared with. Records of data processing activities will be maintained. The aforementioned actions will allow the consortium to comply with the GDPR's accountability principle, which requires organisations to be able to show how they comply with the data protection principles, for example by having effective policies and procedures in place.

Step 3: Communicating privacy information. The HARMONY plan for providing privacy notices already considers the GDPR guidelines. Users who will participate in the pilot surveys will be provided with all needed information, including the project's identity and how we intend to use the collected information through privacy notices. End-users will also be informed of the lawful basis for processing the data, the data retention period and that they have a right to complain to HARMONY if they think there is a problem with the way we are handling their data. All related information will be communicated to end-users in concise, easy to understand and clear language.

Step 4: Individuals' rights. The HARMONY consortium will provide procedures to cover all the rights individuals have, including how personal data are deleted as well as provide data electronically and in a commonly used format. More specifically, the following rights for individuals are considered:

- the right to be informed;
- the right of access;
- the right to rectification;
- the right to erasure;
- the right to restrict processing;
- the right to data portability;
- the right to object; and
- the right not to be subject to automated decision-making including profiling.

Step 5: Subject access requests. Handling data access requests in HARMONY considers the following points:

- No charging will apply for complying with a request.
- Data access requests will be handled within a maximum period of one month.
- The project will refuse requests that are manifestly unfounded or excessive.
- If a request is refused, a clear justification will be provided and which will also inform the individual of the right to complain to the supervisory authority and to a judicial remedy. Any justification will be provided within a maximum period of one month.

² <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr>

Step 6: Lawful basis for processing personal data. A lawful basis for data processing activities has been established and relies on informed consent and privacy notices.

Step 7: Consent. The informed consent forms which will be provided to end-users will comply and meet the GDPR standard. The consent will be freely given, specific, informed and unambiguous. Moreover, it will be separate from other terms and conditions, and will provide simple ways for users to withdraw consent.

Step 8: Children. Although we do not expect underage pilot participants, the age of the users will be verified and parental or guardian consent for any data processing activity will be obtained for underage users.

Step 9: Data breaches. HARMONY establishes procedures to detect, report and investigate a personal data breach. Where a breach is likely to result in a high risk to the rights and freedoms of individuals, these individuals will be notified directly.

Step 10: Data Protection by Design and Data Protection Impact Assessments. HARMONY implements a privacy by design approach. We have already defined WP12 which handles all related aspects.

Step 11: Data Protection Officers. The responsibility for data protection compliance falls under the Data Protection Officers of the partners involved in sensitive data handling, who have the knowledge, support and authority to ensure that the project, its procedures and outcomes adhere to GDPR.

6 Ethical Aspects

Given that HARMONY looks to involve citizens in travel surveys, it is necessary that a governance and ethics framework is embedded within the project. Ethical aspects related to the activities of the project will be managed within WP12 “Ethics requirements”. This work package establishes an effective ethical management, rooted in the project, with a thorough understanding of both the underlying science as well as the associated ethical principles. It covers the management of the project ethical issues related to user studies ensuring the adherence to corresponding regulations. It also includes the provision of consent forms, information sheets and anonymity to participants in the different surveys, while it foresees the monitoring of data sharing frameworks, privacy laws and information law.

6.1 Informed Consent

Participation of persons will be entirely voluntary and we will obtain (and clearly document) their informed consent in advance of their involvement in the HARMONY project. The informed consent form with information sheets will be in a language and in terms fully understandable to participants, describing the aims, methods and implications of the research, the nature of the participation, the amount and nature of the data being stored, any benefits, risks or discomfort that might be involved and the nature of any resulting dissemination. Consent forms will explicitly state that participation is voluntary and that anyone has the right to refuse to participate and to withdraw their participation, samples or data at any time, without any consequences. We will indicate what procedures will be implemented in the event of unexpected or incidental findings. We will ensure that the potential participant has fully understood the information and does not feel pressured or forced to give written consent. Templates of the informed consent/assent forms and information sheets covering the voluntary participation and data protection issues (in language and terms intelligible to the participants), similar to the one submitted in the proposal, will be kept on file and submitted to the Agency upon request.

6.2 Exchanging, archiving and preservation of data

The consortium, within its competences and available infrastructure, will assure secure storage, delivery and access of personal information, as well as managing the rights of the users. In this way, there is complete guarantee that the accessed, delivered, stored and transmitted content will be managed by the right persons, with well-defined rights, at the right time. State-of-the-art firewalls, network security, encryption and authentication will be used to protect collected data (specific details will be developed in the course of the project, within WP3). Firewalls prevent the connection to open network ports, and exchange of data will be through consortium known ports, protected via IP filtering and password.

Where possible (depending on the facilities of each partner) the data will be stored in a locked server, and all identification data will be stored separately. Intrusion Detection systems will monitor anomalies in network traffic and activate restraint policy if needed. A metadata framework will be used to identify the data types, owners and allowable use.

This will be combined with a controlled access mechanism and in the case of wireless data transmission with efficient encoding and encryption mechanisms. Data security will be implemented across all the research sites, and will cover procedures for storage, encryption and transmission of personal data in addition to any national data protection legislation.

The collected data will be stored in a secure server, only visible to the research site network. Anonymous and identifiable data will be stored separately, and only the project authorized person(s) will have access to the stored data. Anonymity will be guaranteed by separating identifiable data from anonymous data. Anonymous data will be available to researchers. If any identifiable data is required for the research purposes, access and distribution to it will be granted only after explicit permission and after agreement of the data holders (participants providing the data). Authentication will be required to access stored data on the research site.

Authorized researchers will have access to the recorded anonymous data after authentication with a centralized server and on a need-to-know basis. Researchers will have access rights to add data to the identity database. No editing or reading rights will be granted to them to prevent alteration/disclosure of private data, if a specific permission is not granted by the data holder.

Those researchers handling and processing personal and sensitive data within the project will be asked to sign a statement that they are familiar with and abide by the contractual obligations of the consortium. If not included in this obligation, they will sign a statement that commits them to make sure project data are not provided to persons outside the project consortium.

When conducting research with vulnerable people and groups honouring and protecting anonymity and confidentiality is especially important. Potential physical, emotional and social dangers to which participants could be exposed through participation will be highly discussed and taken into account. The project consortium will ensure the avoidance of inadvertent reinforcement of negative social stereotypes concerning particular groups and unfair exploitation of vulnerable research participants.

A Data Protection Officer (DPO) will be appointed by all partners involved in personal data handling and the contact details of the DPOs will be made available to all data subjects involved in the research.

As part of follow-up activities and for the preparation of the next version of this initial DMP long term data preservation mechanisms will be explored. Our intention is to preserve non-sensitive data for a predetermined period after the completion of the project. Furthermore, we will identify appropriate archiving institutions that might serve as long term data preservation entities, so that the data produced as part of HARMONY are accessible by the research community in the long term.

7 Conclusions

The Data Management Plan (DMP) presented in the preceding sections outlined the practices that the HARMONY consortium will follow in order to manage data and publications to be used, or to result from the project. This live document will be constantly updated during the project and two further versions (2 and 3) will be produced at M30 and M42 respectively.

In this version of the deliverable we have described the different types of data that we envisage to collect, their sources, the purposes for collecting such data and our planned activities for fair usage of the HARMONY data by the research community. In addition, we have discussed data security and ethical considerations, as well as the resources available for managing data as part of the project.

8 APPENDIX I: Secondary data definition template

Land Use-Transport Interaction Model Data							
1. Contact Details of Representatives							
Full Name:							
Organisation/Department:							
E-mail:							
2. Existing Software and Model Information - if any							
Do you have any operational land use model for your metropolitan area? (if no then proceed to section 3)							
Simulation Software Type (e.g. URBIS XL, LUISA, URBANSIM, TRANUS, MEPLAN, ELUTE, etc.):							
Model spatial coverage (e.g. boundary area or screen shot of the area):							
Model/Version:							
Number of Zones:							
Version date:							
3. Data Requirements Specification – Please, provide information regarding the availability of the requested data types and further descriptions							
Data Types	Can data be used in the project? (Yes/No)	When data will be available to the consortium (Indicate project month e.g. M4: M1 is May 2019)?	Would you like to collect this data during the project? (Yes/No)	Data Format and source? (e.g., csv, text file, ASCII, csv, shapefile...)	Data Year (e.g. 2019)	Description (e.g. Level of Disaggregation - Temporal/Spatial resolution) please provide (if available) maximum disaggregation (e.g. by zone)	Additional Comments
Economic							
Employment						e.g. disaggregation by SIC (Industry type), by occupation, by naga, by floorspace - if by SIC then all data to be indexed by public in time; if by occupation then more than one public in time to be given; if by floorspace then location reference is required	
Retail Activities						e.g. Disaggregation by floorspace, retail sales, employment, sales/expenditure flows - if floorspace then data to be indexed by location, if retail sales then different referents to be given from points to areas to networks	
Freight Flows						e.g. By Industry/Employment Class	
Land use							
Land Use Data						e.g. by type of land use, e.g. industrial, commercial, office, residential; correlated with activity data	
Demographics							
Population						e.g. Usual resident population, resident population in employment, usual workplace population	
Population Projections						e.g. Estimated future resident population	
Occupational Class						e.g. NSOC groups 1-9 (Managers, Professional, Assoc. Prof.,...)	
Ethnicity						e.g. ethnic groups, social classes, ethnogenetic-Latinoamerican, Latin American	
Income						e.g. Mean Income, median income, income percentiles	
Urban density						usually GIS based	
Firm data by industry						e.g. By firm level	
Housing							
Housing Tenure						e.g. Housing tenure (public/private/other) by type (owner/rental) with other variables (e.g. rent, etc.)	
House Prices						e.g. Average Price Paid	
Travel							
Car Ownership						e.g. Number cars per household	
Journey to Work Mode						e.g. when commuting mode percentages - car, taxi, train, metro, walk, etc.	
Commuting Flows Matrix						e.g. residential or workplace commuting flows, meaning disaggregation by mode of transport and by occupational class	
Trip Distributions						e.g. Disaggregation by mode or hour of the day	
Mobile Phonecalls						e.g. related to trip making	
Topography							
General topography						Urban areas/land/coastline geography; rivers and waterbodies	
Point of Interest Data						e.g. from Google Maps, National Mapping Agencies	
Digital Elevation Model						original elevation models; original surface models (contouring, ...)	
Administrative Boundaries							
Municipal Boundaries						e.g. Municipality/Metropolitan area Boundaries	
Local Authority Boundaries						e.g. County/District boundaries	
Transport Infrastructure							
Road network						Road network geography including road class (motorways, ...)	
Walking network						e.g. walk paths, parks, etc.	
Metro Network						Metro lines and stations geography	
Rail Network						Rail lines and stations; high speed rail lines and stations.	
Bus Stops						Bus stop locations	
Airports						Airport locations	
Building Data							
Building Footprints						Building footprint outlines	
Building Floorspace / Heights / Storays						building density aspects: commercial floorspace or building heights, residential, etc.	
Building Function						Function data: office, retail, industrial, residential	
State / Public Housing Estates						Location of large public housing estates	
Planning Policy Data							
Locations for major future urban development						New town locations; opportunity areas; areas for densification	
Planning zoning						zoning restrictions on urban development; existing zoning	
Environmental or other development restrictions						green belt restrictions; national/preserve restrictions; floodline development restrictions	
Major future transport infrastructure development						Major rail, metro and road planned developments	
Other data							
Utility Networks						e.g. water, electricity, waste, telecoms, river	

Organisation:
E-mail:

2. Existing Software and Network Model Information

Scope (e.g. Passenger and/or Freight):

Simulation Software Type (e.g. AIMSUN, VISUM, VISSIM, EMME, SUMO, PARAMICS, TRANSYT, etc.):

Network model spatial coverage (e.g. boundary area or screen shot of the area):

Network Model Baseyear:
Traffic Analysis Zones (in study area):

Traffic Analysis Zones (external):
Number of OD pairs:
Vehicle classes (car, truck, bus, rail, bike, etc.):
License costs:

3. Data Requirements Specification

Please, provide information regarding the availability of the requested data types and further descriptions

Data Type	Can data be used in the project? (Yes/No)	When data will be available to the consortium (indicate project month e.g. M4; M1 is May 2019)?	Would you like to collect this data during the project? (Yes/No)	Data Format and source? (e.g., oem, text file, ASCII, csv, shapefile...)	Data Year (e.g. 2019)	Description (e.g. Level of Disaggregation - Temporal/Spatial resolution/scale)	Additional Comments
Network data							
GIS shapefile of TAZ system and georeferences of centroids and connectors						e.g. OD matrix scales: (number of OD pairs, number of Origin centroids, number of Destination centroids, year)	
GIS shapefile of study area						GIS shapefile content: Information (e.g. does contain lane details, directions and lane-to-lane connectivity in nodes, lane connection information crossings information)	
Open Street Map data						Open street Map data (e.g. does contain lane details, directions and lane-to-lane connectivity in nodes, crossing information)	
Aerial photography and/or CAD for refinements (Google Streetview can be a replacement for this requirement)						Aerial images of the network (potentially Google aerial images can substitute if not available)	
Bike lane network Intersection (node) coordinates						Required	
Section: Road category, speed limit						Required	
Number of lanes						Required	
Length of turn bays						Required	
Lane drop locations						Required	
Lane add locations						Required	
Lane connection information						Required	
Lane channelization						Required	
Link free-flow speed						Required	
Link slope/Grade						Optional	
Lane widths						Optional	Available in google maps
Curvature data						Optional	
Track info in network geometry. See section NETWORK GEOMETRY						e.g. link accessibility for freight vehicles (by vehicle class)	
Location of multimodal transshipment terminals						e.g. transshipment terminals in maritime ports, road-rail terminals, road-barge terminals, etc.	
Location of distribution centers						e.g. location of distribution centers for local or regional distribution channels	

Traffic Control Data		
Sign data (e.g., location of sign data, stop, yield, exit signing and/or lane turning assignment)		
Fixed Signal Control data		fixed control plans, traffic lights position, phases and groups, traffic control plan
Signalized intersections (for each signal groups, signal phases, control plans, coordinates)		
(Semi) Actuated Signal Control data		Data to correctly code controllers (for adaptive signals): SCOOT, MOVA, etc. + turn movement counts + the details of each phase must be obtained, such as minimum green time, maximum green time, yellow change interval time, and red clearance interval time.
Loop detector data		
Ramp Meter Control Data		the metering rate (or headway) associated with a ramp meter and how the metering rate is determined (fixed, ALINEA, HERO, etc.)
Location of variable message signs and set of possible pre-fixed messages		
Location of variable speed signs, operational rules and algorithms used		
Public Transport Data		
GTFS files availability		
Line routing (ideally GIS based)		
Stops location (ideally GIS based)		
Stops assigned to each line		
Timetable covering the simulation period/day		
Type of vehicles used to operate each line (provide fleet size and composition including length, number of doors and which are used for boarding, number of seats, passenger capacity)		
Public transport data on vehicle positions (AVL) – what resolution is it available? Offline or real-time?		
Signal Priority scheme		
Parking Data		
GIS files availability		
On-street parking areas, parking regulation and percentage of occupancy for the period/day to be simulated		
Parking space availability system		
Priority lanes, lane closures for parking during time-of-day/type-of-day, lane or turning closures		
Freight-related data		
City constraints		Regulations imposed by local governments (e.g. forbidding trucks on specific times, vehicle weight restrictions)
Locations for trans-shipment		Transfer locations (intermodal facilities)
Terminal locations		Required
Customer locations		Required
Energy, emission, noise data		
Vehicle Engine Type data		
Noise emission data	Text file	Leq noise level data
Building height	Text file	Height of all structural characteristics e.g. buildings
Type of road surface	Text file	Type and condition of road surfaces: Smooth road, rough etc.
Classification of vehicles	Text file	Category 1: Light motor vehicles, Category 2: Medium heavy vehicles, Category 3: Heavy vehicles, Category 4: Powered two wheelers

Data requirements for Transport Demand Modelling
1. Contact Details of Representative(s)

 Full Name:
 Organisation:
 E-mail:

2. Existing Software and Network Model Information

 Scope (Passenger and/or Freight):
 Simulation Software Type (e.g. AIMSUN, VISUM, VISSIM, SUMO, PARAMICS, TRANSYT, MATSim, etc.):
 4-step or activity-based demand model:
 Study area:
 Baseyear:
 Traffic Analysis Zones (in study area):
 Number of OD pairs:
 Modes (car, truck, bus, rail, bike, etc.):
 License costs:

3. Data Requirements Specification – Please, provide information regarding the availability of the requested data types and further descriptions

Data Type	Can data be used in the project? (Yes/No)	When data will be available to the consortium (indicate project month e.g. M4; M1 is May 2019)?	Would you like to collect this data during the project? (Yes/No)	Data Format and source? (e.g., omx, text file, ASCII, csv, shapefile...)	Data Year (e.g. 2019)	Description (e.g. Level of Disaggregation - Temporal/Spatial resolution/scale)	Additional Comments
Passenger Demand Data							
Household travel demand surveys						e.g. generic household travel demand surveys, most of which include vehicle and parking availability, trips, mode choice, usage and other relevant information	
SP experiments (mode, route or vehicle purchase choice)						Any available market research or stated-preference experiment about mode, route or other travel related choice	
GPS or other geolocation data survey	<input type="checkbox"/>					Detailed trip, activity or time-use diary of individuals containing a typical day or	
Trip or activity or time-use diaries	<input type="checkbox"/>					e.g. OD matrix data: Time Period (e.g. 24 hour, AM and PM peak hour, 07:00-10:00, etc., working day, weekend, year)	
Static OD matrix						e.g. OD matrix data: vehicle type (car, heavy truck, taxi, light truck)	
Static OD matrix per vehicle type							
Data used for static OD matrix estimation and calibration							
Skim matrices						e.g. A skim matrix provides: travel time, distance, costs, or a combination (Generalized Costs), per vehicle type (travel for single-occupancy vehicles, shared-ride 2 and shared-ride 3+, etc.)	
Dynamic OD matrix						e.g. OD matrix data: Time Period (e.g. 15min over 24 hour, 07:00-10:00 with 5, 10, 15 min departure times, etc.,	
Dynamic OD matrix per vehicle type						e.g. OD matrix data: vehicle type (car, heavy truck, taxi, light truck)	
Public transport Data on passenger flows (via on-board counts APC or fare collection AFC) – for what sample? When do people validate? Offline or real-time?							
Pedestrian Data							
Pedestrian counts at crossings per direction (estimated or observed)							
Bicycle counts or occupancy of exclusive bicycle lanes (in case of sharing lanes with other modes provide occupancy share)							
Freight Demand Data							
Freight demand surveys						e.g. Shipper surveys, Firm level surveys	
Truck trip diaries						e.g. Truck trip diaries and observations from EU regions (France, Spain, Italy, Portugal) or outside EU (e.g. UK, Norway, etc.)	
Freight OD matrix						e.g. OD matrix data: by goods type, mode, and/or vehicle type	
Freight OD matrix per mode and/or vehicle type							
Data used for static OD matrix estimation and calibration						e.g. loop detector data by length or weight class	
Skim matrices						e.g. A skim matrix provides: travel time, distance, costs, or a combination (Generalized	
Other Data							

Data requirements for Transport Demand Modelling

1. Contact Details of Representative(s)

Full Name:
Organization:
E-mail:

2. Existing Software and Network Model Information

Scope (Passenger and/or Freight):
Simulation Software Type (e.g. AIMSUN, VISSIM, VISSIM, SUMO, PARAMICS, TRANSYT, MATSim, etc.):
4-step or activity-based demand model:
Study area:
Baseyear:
Traffic Analysis Zones (in study area):
Number of OD pairs:
Modes (car, truck, bus, rail, bike, etc.):
License costs:

3. Data Requirements Specification - Please, provide information regarding the availability of the requested data types and further descriptions

Data Type	Can data be used in the project? (Yes/No)	When data will be available to the consortium (indicate project month e.g. M4, M1 is May 2019)?	Would you like to collect this data during the project? (Yes/No)	Data Format and source? (e.g., oxml, text file, ASCII, csv, shapefile...)	Data Year (e.g. 2019)	Description (e.g. Level of Disaggregation - Temporal/Spatial resolution/scale)	Additional Comments
Passenger Demand Data							
Household travel demand surveys						e.g. generic household travel demand surveys, most of which include vehicle and parking availability, trips, mode choice, usage and other relevant information	
SP experiments (mode, route or vehicle purchase choice)						Any available market research or stated-preference experiment about mode, route or other travel related choice	
GPS or other geolocation data survey	<input type="checkbox"/>					Detailed trip, activity or time-use diary of individuals containing a typical day or	
Trip or activity or time-use diaries	<input type="checkbox"/>					e.g. OD matrix data: Time Period (e.g. 24 hour, AM and PM peak hour, 07:00-18:00, etc., working day, weekend, year)	
Static OD matrix						e.g. OD matrix data: vehicle type (car, heavy truck, taxi, light truck)	
Static OD matrix per vehicle type							
Data used for static OD matrix estimation and calibration						e.g. A skim matrix provides: travel time, distance, costs, or a combination (Generalized Costs), per vehicle type (travel for single-occupancy vehicles, shared-ride 2 and shared-ride 3+, etc.)	
Skim matrices						e.g. OD matrix data: Time Period (e.g. 15min over 24 hour, 07:00-18:00 with 5, 15, 15 min departure times, etc.,)	
Dynamic OD matrix						e.g. OD matrix data: vehicle type (car, heavy truck, taxi, light truck)	
Dynamic OD matrix per vehicle type							
Public transport data on passenger flows (via on-board counts APC or fare collection AFC) - for what sample? When do people validate? Offline or real-time?							
Pedestrian Data							
Pedestrian counts at crossings per direction (estimated or observed)							
Bicycle counts or occupancy at exclusive bicycle lanes (in case of sharing lanes with other modes provide occupancy share)							
Freight Demand Data							
Freight demand surveys						e.g. Shipment surveys, firm level surveys (e.g. truck, rail, carrier and combination modes)	
Truck trip diaries						e.g. OD matrix data: Time Period (e.g. 24 hour, AM and PM peak hour, 07:00-18:00, etc., working day, weekend, year)	
Freight OD matrix						e.g. OD matrix data: by goods type, mode and/or vehicle type	
Freight OD matrix per mode and/or vehicle type						e.g. loop detector data, by length or weight class	
Data used for static OD matrix estimation and calibration						e.g. a skim matrix provides: travel time, distance, costs, or a combination (Generalized Costs)	
Skim matrices							
Other Data							

Model Calibration Data							
1. Contact Details of Representative(s)							
Full Name:							
Organisation:							
E-mail:							
2. Data Requirements Specification - Please, provide information regarding the availability of the requested data types and further descriptions							
Data Type	Can data be used in the project? (Yes/No)	When data will be available to the consortium (Indicate project month e.g. M4; M1 is May 2019)?	Would you like to collect this data during the project? (Yes/No)	Data Format and source? (e.g., omx, text file, ASCII, csv, shapefile...)	Data Year (e.g. 2019)	Description (e.g. Level of Disaggregation - Temporal/Spatial resolution/scale)	Additional Comments
Calibration Data							
Loop detector data						observation of: traffic flow, speed, occupancy, ...?	
Loop detector data						aggregation interval: 5 min	
Loop detector data						time period: 2017, every day	
Loop detector data						data status: raw data or processed data	
Loop detector location data						number of loop detectors	
Travel time for routes with details of the measurement sampling and a clear description (ideally GIS base) of the routes							
Automatic vehicle identification (AVI) data (e.g., cameras, bluetooth stations) - penetration rate, location and semantics of data collection, e.g. aggregation time, offline or real-time availability							
Automatic vehicle location (AVL) data (e.g., Floating car data, GPS, GSM, etc.) - penetration rate, location and semantics of data collection, e.g. aggregation time, offline or real-time availability							
Saturation flow data							
Delay and queue data							
Zone to zone total travel times/waiting times/speed data							
Other							

Mobility Service Data

1. Contact Details of Representative(s)

Full Name:
 Organisation:
 E-mail:

2. Data Requirements Specification - Please, provide information regarding the availability of the requested data types and further descriptions

Data Type	Can data be used in the project? (Yes/No)	When data will be available to the consortium (indicate project month e.g. M4; M1 is May 2019)?	Would you like to collect this data during the project? (Yes/No)	Data Format and source? (e.g., oas, text file, ASCII, csv, shapefile...)	Data Year (e.g 2019)	Description (e.g. Level of Disaggregation - Temporal/Spatial resolution/scale)	Additional Comments
Traditional Taxis (Cabs) - Supply data							
Fleet size: Fleet composition/Vehicle Types/Number for each type: Vehicle Capacities: Fuel Consumption: Others ...							
Ridehailing service data (e.g. Uber, Lyft, Kapten, ViaVan, etc.) - Supply data							
Fleet sizes: Fleet composition/Vehicle Types/Number for each type: Fuel Consumption: Vehicle Capacities: Others ...							
Station-based or free-floating Carsharing (DriveNow, car2go, etc.) - Supply data							
Fleet sizes: Fleet composition/Vehicle Types/Number for each type: Vehicle Capacities: Station capacity/dock-parking numbers: Fuel Consumption: Others ...							
Station-based or free-floating Bikesharing (Santander bikes, Lime, Ofo, OBiike, etc.) -Supply data							
Fleet sizes: Fleet composition/Vehicle Types/Number for each type: Vehicle Capacities: Station capacity/dock-parking numbers: Fuel Consumption: Others ...							
Station-based or free-floating Scooters (Lime, Voi, etc.) - Supply data							
Fleet sizes: Fleet composition/Vehicle Types/Number for each type: Vehicle Capacities: Station capacity/dock-parking numbers: Fuel Consumption: Others ...							
Ride-sourcing/Ride-hailing/E-hailing/Ridesharing apps (Uber, Gett, Kapten, ViaVan, etc.) - Demand data							
Trips /ODs: Driving Profiles: Others:							
Station-based or free-floating Carsharing (DriveNow, car2go, etc.)							
Static or dynamic daily station/zone stock: Level data: Trips /ODs: Rentals/Bookings: Others:							

<ul style="list-style-type: none"> Fleet sizes Fleet composition/Vehicle Types/Number for each type Vehicle Capacities Station capacity/dock-parking numbers Fuel Consumption Others ... 	Station-based or free-floating Scooters (Lime, VOI, etc.) - Supply data
<ul style="list-style-type: none"> Trips /ODs Driving Profiles Others 	Ride-sourcing/Ride-hailing/E-hailing/Ridesharing apps (Uber, Gett, Kapten, ViaVan, etc.) - Demand data
<ul style="list-style-type: none"> Static or dynamic daily station/zone stock level data Trips /ODs Rentals/Bookings Others 	Station-based or free-floating Carsharing (DriveNow, car2go, etc.)
<ul style="list-style-type: none"> Static or dynamic daily station/zone stock level data Trips /ODs Rentals/Bookings Others 	Station-based or free-floating Bikesharing (Santander bikes, Lime, Ofo, OBIke, etc.)
<ul style="list-style-type: none"> Static or dynamic daily station/zone stock level data Trips /ODs Rentals/Bookings Others 	Station-based or free-floating Scooters (Lime, VOI, etc.)
<ul style="list-style-type: none"> Fleet sizes Fleet composition/Vehicle Types/Number for each type Vehicle Capacities Speed Fuel consumption Zone to zone travel times and loading/unloading times Others 	Traditional freight operators (Trucks)
<ul style="list-style-type: none"> Fleet sizes Fleet composition/Vehicle Types/Number for each type Vehicle Capacities Speed Fuel consumption Zone to zone travel times and loading/unloading times Others 	Crowdshipping
<ul style="list-style-type: none"> Fleet sizes Fleet composition/Vehicle Types/Number for each type Vehicle Capacities Speed Charging facilities Zone to zone travel times and loading/unloading times Others 	Cargo Bikes

9 APPENDIX II: Model Outputs data definition template

Data set category (Data provided from the HARMONY models)	
Data set description	
What is the purpose of data collection/generation? "Data utility": to whom this dataset will be useful?	
Data set reference and name	
Who (partner name) / When (which task and when the data will be available)	
Format (including related standards and metadata). For example, you can check whether any standards listed in the Metadata Standards Directory of Research Data Alliance makes sense for your data (http://rd-alliance.github.io/metadata-directory/)	
In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how	
Data sharing plan (license) If open, please describe how it will be made available (e.g. submission to a repository?)	

Relation to project Objective(s) - for which objective is this dataset relevant and why?	
Pre-existing dataset or new? (if pre-existing please provide a reference)	
Size / expected size (use a measurement unit that makes sense for the dataset)	



[@Harmony_H2020](https://twitter.com/Harmony_H2020)

#harmony-h2020



<https://www.linkedin.com/company/harmony-h2020/>

For further information please visit www.harmony-h2020.eu

