



DATA REQUIREMENTS AND AVAILABLE DATA SOURCES

D2.1



This project has received funding from
the European Union's Horizon 2020
research and innovation programme
under grant agreement No 955273

Deliverable administrative information

Deliverable number	D 2.1
Deliverable title	Data requirements and available data sources
Dissemination level	Public
Submission deadline	31/08/2022
Version number	1.0
Authors	Marco Comerio (Cefriel) Antonia Azzini (Cefriel) Mario Scrocca (Cefriel) Andrea Fiano (Cefriel)
Internal reviewers	Ana V. Silva (A-to-Be) Ynte Vanderhoydonc (IMEC) Athina Tympakianaki (Aimsun) Juan S. Angarita (Deusto) Leire Serrano (Deusto) Antonio D. Masegosa (Deusto) Eleni Montouka (NTUA)
Document approval	João Vieira (CARRIS) Hannah Tune (TFGM) Isabelle Dussutour (ID4CAR) Céline Quéron, Pierre Renault (Rennes Metropole) Sébastien Leparoux (Keolis Rennes)

Legal Disclaimer

TANGENT is co-funded by the European Commission, Horizon 2020 research and innovation programme under grant agreement No. 955273 (Innovation Action). The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any specific purpose. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The TANGENT Consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Copyright © TANGENT Consortium, 2021.



https://twitter.com/TANGENT_H2020



<https://www.linkedin.com/company/tangent-project/>



https://www.youtube.com/channel/UCjhz4kwEm_sTHj7fE4zXToA

For further information please visit <http://www.tangent-h2020.eu/>

Executive summary

The gathering of data sources is a fundamental activity to enable the development and testing of the TANGENT services for multimodal transport network management in the Athens, Lisbon, Manchester, and Rennes case studies.

This deliverable provides a list of data requirements for the TANGENT services extracted from the specification of the TANGENT services and case studies reported in D1.2 “NTM needs assessment and system requirements”, and from the specification of the components that are under development by the WP3-6 of the project.

Moreover, this deliverable provides the current list of available data sources to fulfil the identified data requirements. The list has been defined through (i) the creation of data catalogues for the case studies with the involvement of the TANGENT case study leaders, (ii) the usage of the catalogues to identify the data sources that fulfil the data requirements, and (iii) refinement and validation of the mappings between data requirements and the data sources with the project partners.

Finally, this deliverable describes the analysis of the selected data sources with a particular focus on the following aspects to be managed in the next steps of the project:

- current gaps in fulfilling the data requirements;
- accessibility of data and licensing issues;
- data interoperability issues to be addressed by the TANGENT project through developing the WP2 solution for data harmonization and fusion (first release at M20).

Keywords

Data requirements, data sources, data interoperability, metadata, data harmonization.

Table of contents

DELIVERABLE ADMINISTRATIVE INFORMATION.....	1
EXECUTIVE SUMMARY	3
TABLE OF CONTENTS	4
LIST OF ABBREVIATIONS AND ACRONYMS.....	7
1 INTRODUCTION	9
2 DATA REQUIREMENTS.....	11
3 DATA SOURCE COLLECTION	20
4 DATA INTEROPERABILITY	42
5 CONCLUSIONS	50
6 REFERENCES	51
ANNEX I	52
ANNEX II	53
ANNEX III	54
ANNEX IV	55

List of figures

Figure 1: Connections between WP2 and other WPs	10
Figure 2: The methodology for the elicitation of data requirements	11
Figure 3: Extraction of data requirements from TANGENT component descriptions available in the Technical Coordination Spreadsheet.....	12
Figure 4: Graphical representation of the data source collection task.	20
Figure 5: Challenges to implement data interoperability	42
Figure 6: Generic scenario considering different data catalogues and data sources	43
Figure 7: Any-to-any versus any-to-one mapping approaches.....	45
Figure 8: Snapshot of data sources described in the Athens catalogue	52
Figure 9: Snapshot of data sources described in the Lisbon catalogue.....	53
Figure 10: Snapshot of data sources described in the Manchester catalogue	54
Figure 11: Snapshot of data sources described in the Rennes catalogue (Part 1).....	55
Figure 12: Snapshot of data sources described in the Rennes catalogue (Part 2).....	56

List of tables

Table 1: List of data requirements identified for TANGENT services.....	15
Table 2: Priorities for each data requirement considering the TANGENT components to be developed	17
Table 3: Filtered and refined list of data requirements considering the current version TANGENT case studies	19
Table 4: List of potential data sources to be collected in the TANGENT project.	23
Table 5: The TANGENT metadata profile.....	26
Table 6: Example of Bus data sheet of Rennes Data Catalogue	27
Table 7: Available data portals and online data services to identify additional data sources	30
Table 8: Mapping between data requirements and available data sources (LISBON case study)	32
Table 9: Mapping between data requirements and available data sources (MANCHESTER case study)	36
Table 10: Mapping between data requirements and available data sources (RENNES case study) ...	39

Table 11: Mapping between data requirements and available data sources (ATHENS case study) ...41

Table 12: TANGENT data interoperability issues for the LOCATE and ACCESS challenges.....49

List of abbreviations and acronyms

Acronym	Meaning
API	Application Program Interface
ASAP	As Soon As Possible
BRT	Bus Rapid Transit
CAV	Connected and Autonomous Vehicles
CCTV	Closed-circuit television
CIM	Cooperative Incident Management
CMC	Coordinated Metadata Catalogue
COP	Common Operational Picture
DCAT-AP	Data Catalog Application Profile
DOA	Description of Actions
DRT	Demand Responsive Transport
EC	European Commission
EU	European Union
FCD	Flow Car Data
GA	Grant Agreement
GM	Greater Manchester
GMP	Greater Manchester Police
GMSM	Greater Manchester SATURN Model
IoT	Internet of Things
IVI	In-Vehicle Information
KPI	Key Performance Indicator
LoS	Level of Service
NAP	National Access Point
NLB	Network Load Balance
NTM	Network Traffic Management
OD	Origin-Destination
ODP	Open Data Portal
P&R	Park and Ride
PKI	Portuguese Key authentication Infrastructure

PT	Public Transport
PTU	Public Transport Authority
SNLB	Smart Network Load Balance
SRN	Strategic Route Network
UTC	Urban Traffic Control
UTMC	Urban Traffic management and Control
UVAR	Urban Vehicle Access Regulations
VMS	Variable Message Signs
WP	Work Package
w.r.t.	With Respect To

1 Introduction

This chapter introduces the objectives of the deliverable and provides an overview of the results presented.

1.1 Attainment of the objectives and explanation of deviations

This deliverable aims to provide (i) a list of data requirements for the TANGENT services, (ii) available data sources to fulfil the data requirements, and (iii) data interoperability issues to be addressed by the WP2 of the project.

Data requirements have been derived from the specification of the TANGENT services and use cases reported in D1.2 “NTM needs assessment and system requirements,” and from the specification of the components that are under development by WP3-6 of the project.

Data collection has been performed with the direct involvement of the TANGENT case study leaders. In parallel with the data requirement elicitation, a data catalogue for each case study has been produced to have the first map of available data sources. Particular attention has been dedicated to travel and traffic data available on National Access Points, defined according to EU Directive 2010/40 and all its supplementing delegated regulations (EU Delegated Regulations 886/2013, 2015/962, 1926/2017). The produced data catalogues have been then used to identify the data sources that fulfil the data requirements and potential gaps to be managed by the project.

The analysis of the available data sources has guided the identification of (i) current gaps in fulfilling the data requirements; (ii) accessibility of data and licensing issues to be solved through the signature of specific NDA, and (iii) data interoperability issues to be managed by the TANGENT project through the development of the WP2 solution for data harmonization and fusion (first release at M20).

1.2 Intended audience

This deliverable is mainly intended for project partners, the project officers and the public interested in the data gathering process, particularly the case study stakeholders and data providers.

1.3 Structure of the deliverable and links with other work packages/deliverables

Section 2 of this deliverable describes the methodology adopted for data requirement elicitation and provides the list of data requirements for each TANGENT case study.

Then, Section 3 describes all the activities carried out with the direct involvement of case study leaders and stakeholders to identify open and private data sources to be potentially used for testing the TANGENT solution. An analysis of the coverage of the collected data sources w.r.t. the identified data requirements is also provided.

Afterwards, Section 4 reports on data interoperability issues to be solved through the development of the WP2 solution for data harmonization and fusion.

Concluding remarks are reported in Section 5. The data catalogues produced to support the data collection activity for each case study are summarised and reported as Annexes I-IV of this deliverable.

In summary, the deliverable provides the first map of TANGENT data requirements, available data sources, and data interoperability issues to be addressed. The refinement of data requirements and the

related data collection activity is expected to proceed along the project lifecycle to better address the specifications of the TANGENT services and the use cases.

Figure 1 shows the activities performed by Task 2.1, described in this deliverable, together with connections with other TANGENT WPs. Task 2.1 has received inputs from WP1 (i.e., specifications of the TANGENT services and case studies), WP3-6 (i.e., data requirements for the TANGENT components), and WP7 (i.e., data sources availability). The output of Task 2.1 (i.e., this deliverable) will guide the activities of Task 2.3 related to the development of the software solution.

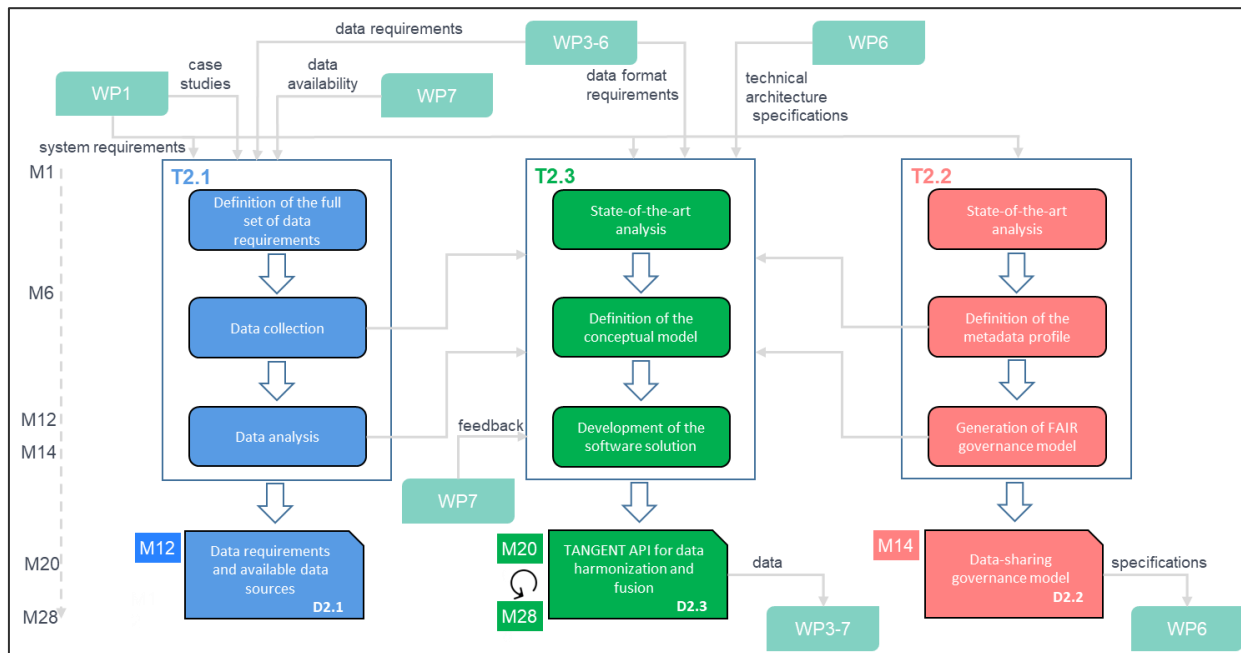


Figure 1: Connections between WP2 and other WPs

2 Data Requirements

Data requirements aim to identify which data is necessary to enable the development and testing of the TANGENT services. Data requirements are specific for each TANGENT case study and are used to guide the data collection activities (see Section 3).

In this section, the methodology applied for the elicitation of data requirements is described together with the first list of requirements identified for each case study.

2.1 The methodology for the elicitation of data requirements

To facilitate the understanding and comprehension of the methodology adopted for the elicitation of data requirements, a short summary of key elements and concepts of the software platform to be developed in TANGENT is provided. The complete description of these elements is available in D1.2 “NTM needs assessment and system requirements.”

- A set of TANGENT services (i.e., Enhanced Information Service for Multimodal Transport Management, Real-Time Traffic Management Services, and Transport Network Optimization for Transport Authorities) has been defined. Each service is characterized by specific functionalities offered to the users.
- Each TANGENT service functionality generates a set of system requirements.
- The system requirements are associated with specific TANGENT components (e.g., WP4-Supply Forecasting, WP6-Dashboard) that are software solutions under design and development by TANGENT WP3-6.
- Each TANGENT component is currently defined in terms of expected inputs and outputs to be produced. The TANGENT component description tables are included in the Technical Coordination Spreadsheet produced by WP6 (e.g., the WP4-Supply Forecasting requires, among others, “live road traffic measurements” to produce “forecasting on road traffic status”).
- The TANGENT services will be tested in specific case studies (Athens, Lisbon, Manchester, and Rennes). The use cases description for each case study provides, among other information, (i) which TANGENT services will be tested and (ii) the expected visualization and interaction.

Figure 2 shows the methodology adopted for the elicitation of data requirements.

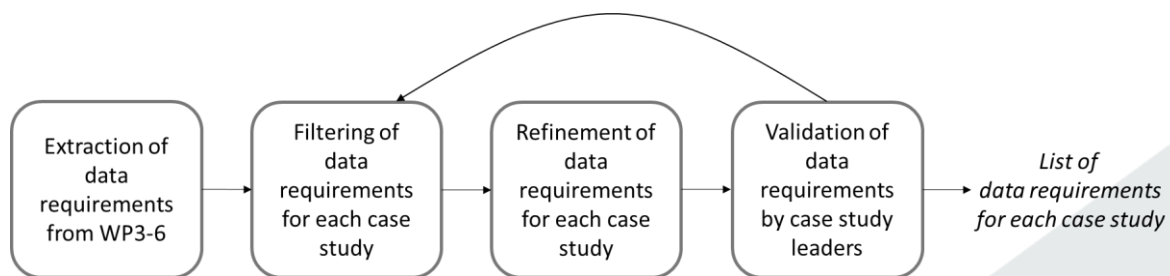


Figure 2: The methodology for the elicitation of data requirements

The first set of data requirements is extracted from the description of the TANGENT components under definition and development by TANGENT WP3-6. As shown in Figure 3, the data item inputs of each component that are expected to be provided by TANGENT WP2 solution become data requirements.

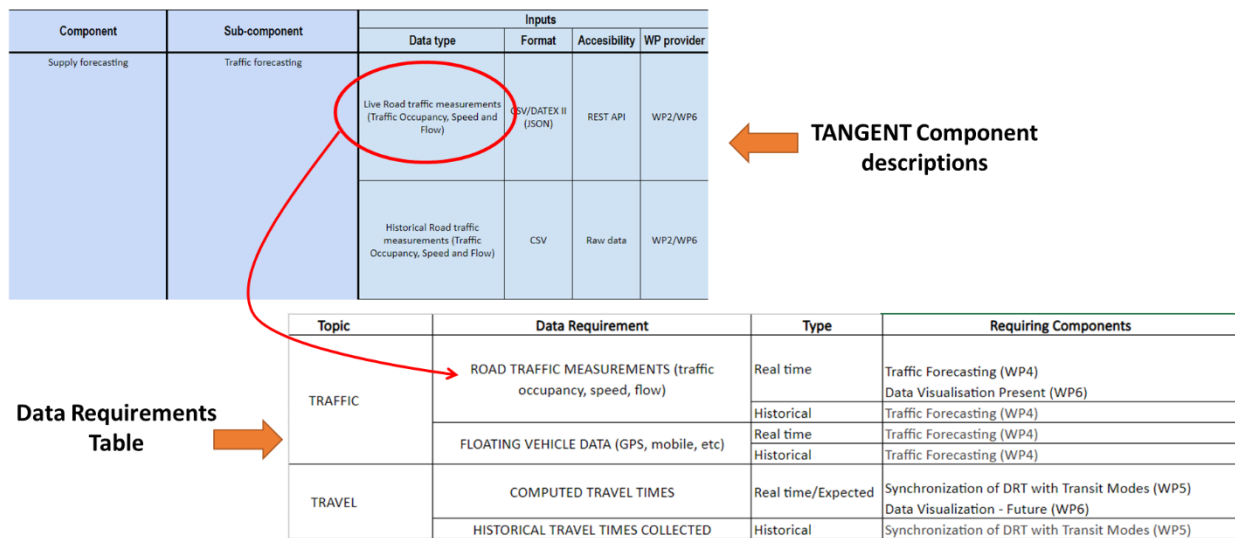


Figure 3: Extraction of data requirements from TANGENT component descriptions available in the Technical Coordination Spreadsheet

The second step of the methodology consists of producing a data requirement table for each case study. As mentioned above, the data requirements are extracted from the description of the TANGENT components that enable the TANGENT services. Therefore, the data requirement table for a specific case study includes only the requirements associated with the services to be tested¹.

The third step of the methodology consists of refining each case study's data requirement table. The use case descriptions of the TANGENT case studies include the tables "Visualization and Interaction" that identify the data expected to be visualized by the users². Moreover, the WP1 workshops with stakeholders, reported in D1.2, supported collecting their needs and expectations in using the TANGENT services. The analysis of these results enables the refinement of each data requirement table with more specific requirements related to the case study.

In the last methodological step, the case study leaders are involved in validating the identified requirements. In this document we provide a first and consolidated version of the collected and validated data requirements. Nevertheless, during the following stages of the project, additional data sources may be required, so the collection and validation of data requirements will be an iterative process.

The results of each step of the methodology will be described in the next sections.

2.2 Data Requirements

This section describes the result of the first step of the methodology, where data requirements are extracted from the descriptions of the TANGENT components. Table 1 represents the first list of data requirements, each associated with the TANGENT components requiring it.

Currently, **36 data requirements** have been identified that can be grouped by topics, as follows:

- 11 data requirements related to road and public TRANSPORT NETWORK;
- 5 data requirements related to ROAD TRANSPORT;

¹ The list of services being tested in each case study is reported in Table 1 of D1.2 "NTM needs assessment and system requirements."

² The reader can refer to Tables 8, 16, 22 and 29 of D1.2 "NTM needs assessment and system requirements."

- 8 data requirements related to PUBLIC TRANSPORT services;
- 9 data requirements related to INFLUENCING FACTORS (e.g., weather, planned events);
- 3 OTHERS data requirements related to DCP policy zone, DRT fleet characteristics and other transportation modes.

It is important to note that the list of data requirements presented in Table 1 does not include requirements associated with some of the components (in particular WP4 and WP5 components) that will be fulfilled by data sources extracted or derived from the calibrated transport models. The types of data sources that will be extracted from the calibrated transport models are:

- OD road travel demand matrix (aggregated, disaggregated)
- Public Transport OD demand matrix (aggregated, disaggregated)
- DRT disaggregated demand
- Public Transport Fleet characteristics
- Traffic control plans

These data requirements are not included in the overall data collection process discussed in this document since the calibrated transport models from which they will be extracted are general requirements for the TANGENT solution and were already identified. The existing transport models, for the different case studies, are listed in the TANGENT Deliverable "D1.2 NTM needs assessment and system requirements".

Topic	Data Requirement		Requiring Components
TRANSPORT NETWORK DATA	ROAD TRANSPORT NETWORK (roads, limited access zones, etc)		Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4) Dashboard - Data visualisation present (WP6)
	ROAD EQUIPMENT POSITION		Data Processing - Current traffic state (WP4) Supply Forecasting – Traffic forecasting (WP4) Dashboard - Data visualisation present (WP6)
	PUBLIC TRANSPORT NETWORK		Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4) Dashboard - Data visualisation present (WP6)
	ROAD TRANSPORT NETWORK EVENTS (planned)	Real time	Critical Conditions - Anomalous event detection (WP4) Dashboard - Data visualisation present (WP6) Monitoring Module (WP6)
		Historical	Critical Conditions - Anomalous event detection (WP4)
	PUBLIC TRANSPORT NETWORK EVENTS (planned)	Real time	Critical Conditions - Anomalous event detection (WP4) Dashboard - Data visualisation present (WP6) Monitoring Module (WP6)
		Historical	Critical Conditions - Anomalous event detection (WP4)
	ROAD TRANSPORT NETWORK	Real time	Critical Conditions - Anomalous event detection (WP4) Dashboard - Data visualisation present (WP6) Monitoring Module (WP6)

Topic	Data Requirement		Requiring Components
	INCIDENTS (unplanned)	Historical	Critical Conditions - Anomalous event detection (WP4)
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)	Real time	Critical Conditions - Anomalous event detection (WP4) Dashboard - Data visualisation present (WP6) Monitoring Module (WP6)
		Historical	Critical Conditions - Anomalous event detection (WP4)
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)	Real time	Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4) Demand Estimation - Demand prediction (WP4) Dashboard - Data visualisation present (WP6)
		Historical	Supply Forecasting - Traffic forecasting (WP4) Demand Estimation - Offline demand estimation (WP4)
	FLOATING VEHICLE DATA (GPS, mobile, etc)	Real time	Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4)
		Historical	Supply Forecasting - Traffic forecasting (WP4) Demand Estimation - Offline demand estimation (WP4)
	ROAD TRAVEL TIMES (external services, statistics, etc)		Data Processing - Current traffic state (WP4) Demand Estimation - Offline demand estimation (WP4) Synchronization of DRT with Transit Modes (WP5) Dashboard - Data visualisation present (WP6)
PUBLIC TRANSPORT	SCHEDULES AND LINES		Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4) Synchronization of DRT with Transit Modes (WP5) Synchronization of Public Transport and Traffic Control (WP5) Dashboard - Data visualisation present (WP6)
	TICKET VALIDATION DATA (e.g., smart card usage)		Demand Estimation - Demand prediction other modes (WP4)
	FLOATING PT VEHICLE DATA (GPS, mobile, etc)	Real time	Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4)
		Historical	Supply Forecasting - Traffic forecasting (WP4)
	DELAYS	Real time	Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4)

Topic	Data Requirement		Requiring Components
	OCCUPANCY	Historical	Supply Forecasting - Traffic forecasting (WP4) Demand Estimation - Offline demand estimation (WP4)
		Real time	Data Processing - Current traffic state (WP4) Supply Forecasting - Traffic forecasting (WP4)
		Historical	Supply Forecasting - Traffic forecasting (WP4)
INFLUENCING FACTORS	WEATHER DATA (measurements, e.g., temperature, humidity, air quality, etc)	Real time	Supply Forecasting - Traffic forecasting (WP4) Dashboard - Data visualisation present (WP6)
		Historical	Supply Forecasting - Traffic forecasting (WP4)
	WEATHER EVENTS	Real time	Supply Forecasting - Traffic forecasting (WP4) Dashboard - Data visualisation present (WP6)
		Historical	Supply Forecasting - Traffic forecasting (WP4)
	FORECASTED WEATHER DATA	Real time	Supply Forecasting - Traffic forecasting (WP4)
		Historical	Supply Forecasting - Traffic forecasting (WP4)
	HOLIDAY/SEASONAL CALENDAR		Supply Forecasting - Traffic forecasting (WP4)
	PLANNED EVENTS (sport, entertainment, etc)	Real time	Supply Forecasting - Traffic forecasting (WP4) Critical Conditions - Anomalous event detection (WP4) Monitoring Module (WP6)
		Historical	Critical Conditions - Anomalous event detection (WP4)
OTHERS	AREA OF DCP POLICY ZONE		Dynamic Congestion Pricing (WP5)
	DRT FLEET CHARACTERISTICS		Synchronization of DRT with Transit Modes (WP5)
	DATA ABOUT OTHER TRANSPORTATION MODES (shared mobility, taxis, etc)		Dashboard - Data visualisation present (WP6) Dashboard – Smart network load balance (WP6) Dashboard – Cooperative incident management (WP6) Monitoring Module (WP6)

Table 1: List of data requirements identified for TANGENT services

To support the identification of relevant gaps in the availability of data sources for each TANGENT case study, we also collected the priority associated with TANGENT components requiring it. Table 2 reports the priority for each data requirement, grouping them by the WP responsible for developing the components.

Topic	Data Requirement	Type	Priority WP4	Priority WP5	Priority WP6
TRANSPORT NETWORK	ROAD TRANSPORT NETWORK (roads, limited access zones, etc)		High	Low	High
	ROAD EQUIPMENT POSITION		High	Low	Low
	PUBLIC TRANSPORT NETWORK		Low	Low	Medium
	ROAD TRANSPORT NETWORK EVENTS (planned)	Real time	High	Not required	High
		Historical	High	Not required	Not required
	PUBLIC TRANSPORT NETWORK EVENTS (planned)	Real time	High	Not required	High
		Historical	High	Not required	Not required
	ROAD TRANSPORT NETWORK INCIDENTS (unplanned)	Real time	High	Not required	High
		Historical	High	Not required	Not required
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)	Real time	High	Not required	High
		Historical	High	Not required	Not required
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)	Real time	High	Not required	High
		Historical	High	Not required	Not required
	FLOATING VEHICLE DATA (GPS, mobile, etc)	Real time	High	Not required	Not required
		Historical	High	Not required	Not required
	ROAD TRAVEL TIMES (external services, statistics, etc)		Medium	Medium	Medium
PUBLIC TRANSPORT	SCHEDULES AND LINES		High	High	High
	TICKET VALIDATION DATA (e.g., smart card usage)		Low	Not required	Not required
	FLOATING PT VEHICLE DATA (GPS, mobile, etc)	Real time	High	Not required	Not required
		Historical	High	Not required	Not required
	DELAYS	Real time	High	Not required	Not required
		Historical	High	Not required	Not required
	OCCUPANCY	Real time	High	Not required	Not required
		Historical	High	Not required	Not required
INFLUENCING FACTORS	WEATHER DATA (measurements, e.g., temperature, humidity, air quality, etc)	Real time	Low	Not required	Low
		Historical	Low	Not required	Not required
	WEATHER EVENTS	Real time	Low	Not required	Medium
		Historical	Low	Not required	Not required

Topic	Data Requirement	Type	Priority WP4	Priority WP5	Priority WP6
	FORECASTED WEATHER DATA	Real time	Low	Not required	Not required
		Historical	Low	Not required	Not required
	HOLIDAY/SEASONAL CALENDAR		Low	Not required	Not required
	PLANNED EVENTS (sport, entertainment, etc.)	Real time	Low	Not required	Medium
		Historical	Low	Not required	Not required
OTHERS	AREA OF DCP POLICY ZONE		Not required	Medium	Not required
	DRT FLEET CHARACTERISTICS		Not required	Low	Not required
	DATA ABOUT OTHER TRANSPORTATION MODES (shared mobility, taxis, etc)		Not required	Low	Medium

Table 2: Priorities for each data requirement considering the TANGENT components to be developed

2.3 Filtering and refining data requirements

As explained above, the methodology's second and third steps consist of producing a data requirement table for each case study thought:

- (I). Filtering out from Table 1 requirements associated with services not to be tested by the case study;
- (II). Refining the requirements focusing on specific characteristics of the case study, based on the inputs from the pilot cities stakeholders.

Table 3 shows how the data requirements in Table 1 are refined for each case study:

- Data requirements related to "DRT Fleet Characteristics" and "Area of DCP Policy Zone" will no longer be considered for the Lisbon, Manchester, and Rennes case studies since they are associated with the TANGENT services "CIM - Synchronization of on-demand and PT services" and "NLB – Dynamic Congestion Pricing" that will be tested only in Athens.
- The data requirements for the case studies are refined according with:
 - the expected type of road traffic measurements;
 - the type of requested floating vehicle data;
 - specific external services providing travel times;
 - the type of requested statistics related to travel time;
 - the type of requested weather data;
 - specific data related to road transport network and equipment;
 - specific events on the transport network to be considered;
 - the considered PT modes.
- Several data requirements do not apply to the Athens case study. Considering that it will be a virtual case study, the data requirements for Athens are mainly associated with data sources needed to (re-)calibrate the Athens transport model.

Topic	Data Requirement	Athens	Lisbon	Manchester	Rennes
TRANSPORT NETWORK	ROAD TRANSPORT NETWORK (roads,	Yes	Yes	Yes	Reserved bus lanes

Topic	Data Requirement	Athens	Lisbon	Manchester	Rennes
	limited access zones, etc)				
	ROAD EQUIPMENT POSITION	Traffic loops	Traffic lights, Traffic loops, Cameras, Other sensors	CCTV, Sensors, Traffic Signals, VMS	Traffic lights, Traffic loops
	PUBLIC TRANSPORT NETWORK	Bus, Metro	Stops for Bus, Tram, Train, Metro	Bus and Tram	Stops for Bus, Tram, Train, Metro
	ROAD TRANSPORT NETWORK EVENTS (planned)	Yes	Positions, start and duration of the event	Positions, start and duration of the event	Positions, start and duration of the event
	PUBLIC TRANSPORT NETWORK EVENTS (planned)	No	Yes	Yes	Positions, start and duration of the event
	ROAD TRANSPORT NETWORK INCIDENTS (unplanned)	Yes	Positions, start and duration of the event	Positions, start and duration of the event	Positions, start and duration of the event
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)	Yes	Yes	Yes	Positions, start and duration of the event
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)	Yes	Traffic counters, Tolling transaction	Traffic counters	Traffic flows, Congestion of Reserved Bus Lanes
	FLOATING VEHICLE DATA (GPS, mobile, etc)	No	C-ITS data	Yes	Floating Car Data (FCD)
	ROAD TRAVEL TIMES (external services, statistics, etc)	No	Waze, TomTom, Google Maps, etc. for Journey time	Journey time for routes	Floating Car Data (FCD)
PUBLIC TRANSPORT	SCHEDULES AND LINES	Bus	Bus, Tram, Train, Metro	Bus and Tram	Bus and Metro
	TICKET VALIDATION DATA (e.g., smart card usage)	Yes	Yes	Yes	Bus and Metro
	FLOATING PT VEHICLE DATA (GPS, mobile, etc)	No	Bus and Tram	Bus	Bus
	DELAYS	No	Yes	Tram	Bus and Metro
	OCCUPANCY	No	Bus and Tram	Tram	Bus and Metro

Topic	Data Requirement	Athens	Lisbon	Manchester	Rennes
					(ticketing data)
INFLUENCING FACTORS	WEATHER DATA (measurements, e.g., temperature, humidity, etc)	No	Yes	Yes	Weather information and air quality data
	WEATHER EVENTS	No	Yes	Yes	Yes
	FORECASTED WEATHER DATA	No	Yes	Yes	Yes
	HOLIDAY/SEASONAL CALENDAR	No	Yes	Yes	National education minister holiday calendar
	PLANNED EVENTS (sport, entertainment, etc)	Large events at the outskirts of Athens	Major events (Web Summit, Jornadas da Juventude, Rock in Rio, NOS Alive), sport events (major football match), cultural events (Santos Populares)	Concerts, football matches, public transport strike	Specific events, SPACE Event, Stade Rennais matches
OTHER	AREA OF DCP POLICY ZONE	Yes	No	No	No
	DRT FLEET CHARACTERISTICS	Yes	No	No	No
	DATA ABOUT OTHER TRANSPORTATION MODES (shared mobility, taxis, etc)	No	Yes	Yes	Yes

Table 3: Filtered and refined list of data requirements considering the current version TANGENT case studies

3 Data Source Collection

3.1 Approach and Objectives

The data collection activity aims to identify data sources available for the four TANGENT case studies. In TANGENT, each case study city defines the mobility and transport context independently, highlighting the transport modes peculiar to the project and other relevant information.

The available data sources for the various transport modes are diverse and often defined through different formats. For this reason, in conjunction with the definition of the data requirements outlined in Section 2, a TANGENT catalogue is created to have a broader view of the various types of data available for each case study.

At the moment of writing this deliverable, each catalogue is an excel document consisting of:

- two sheets with guidelines to compile the catalogue and general information on the case study;
- a sheet for each transport mode with the metadata description of the identified data sources;
- the last sheet with contextual data sources useful for the project (e.g., environmental data).

The TANGENT catalogues are not currently in their final form; indeed, the final version of the TANGENT catalogues will be an instance of a metadata catalogue, a digital platform supporting the sharing, reuse, and governance of distributed digital assets. This metadata catalogue will be customized according to the TANGENT metadata profile and data sharing governance that will be described in D.2.2 “Data Sharing Governance Model”.

Figure 4 describes the different phases of the data collection task. The goal is to release a tool that collects data sources described according to a metadata profile common to all the case studies.

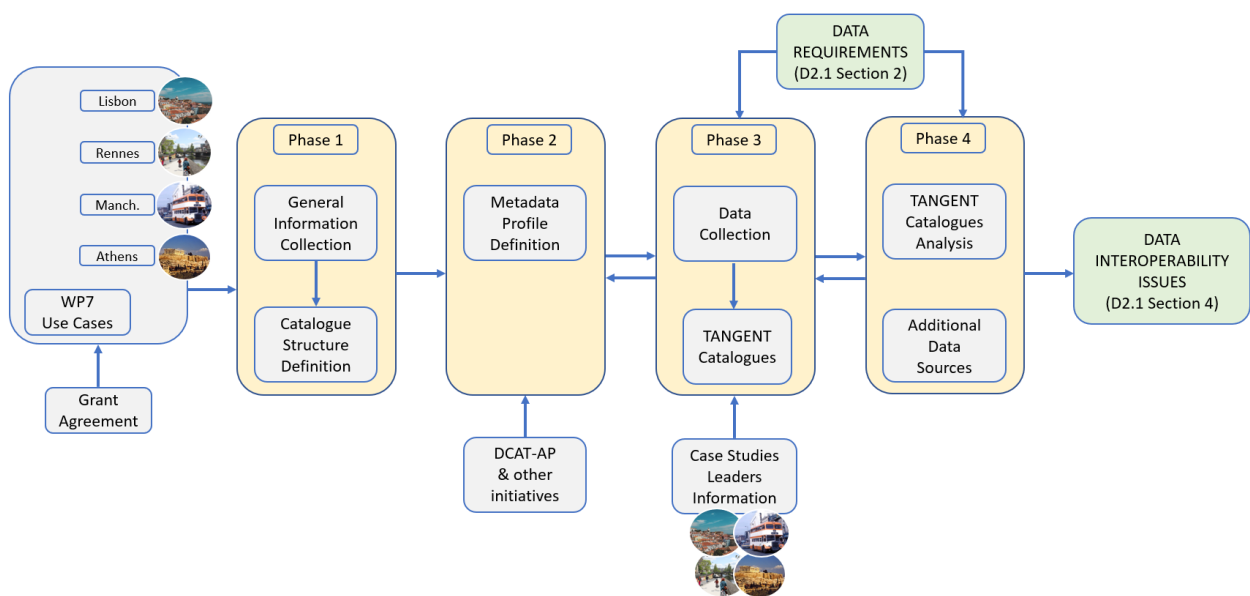


Figure 4: Graphical representation of the data source collection task.

The phases of the data collection task are:

- Phase 1: General Information Collection & Catalogue Structure Definition: in this initial phase, a set of guidelines for compiling the catalogue is provided with general information on the case study.
- Phase 2: Metadata Profile Definition: different data sources must be described according to specific metadata. The activity carried out in this phase consists of the definition of a reference metadata profile to describe all data sources homogeneously and uniformly.
- Phase 3: Data Source Collection: gathering data sources requires continuous interaction among the TANGENT partners. In particular, WP2 is involved in frequent interactions, still in progress, with case study leaders and their stakeholders. The aim is to obtain a catalogue of data sources as complete as possible to support strategic decisions to be taken for subsequent project tasks. A clear definition of the catalogue structure and the adopted metadata profile is needed to facilitate data collection and to provide a comprehensive description of the data sources. For this reason, iterations between Phase 2 and Phase 3 are necessary.
- Phase 4: TANGENT Catalogues Analysis: The data sources described in the catalogue are analyzed in this phase. The analysis aims to assess the availability of data sources to cover the data requirements identified for the case study (see Section 2). Furthermore, an analysis of available open data catalogues and data services is performed to identify additional data sources that could be used to fill gaps in the mapping between data requirements and data sources available in the catalogues.

The overall goal of the data collection is to create, for each case study, a catalogue of available data sources to facilitate the mappings between these data sources and the data requirements throughout the project lifecycle.

The TANGENT catalogues must be considered the first consolidated release of the data collection activity. However, during the following stages of the project, additional data requirements may arise, requiring further data collection and the update of the data catalogues.

3.2 TANGENT Metadata Catalogues

This subsection provides a detailed description of each phase of the data collection activity.

3.2.1 Phase 1: The structure of the catalogues and general information

The first objective of Phase 1 is to provide a high-level description of the transport systems of each TANGENT case study to guide the data collection task.

Starting from the TANGENT Grant Agreement, a set of general information defined for each TANGENT case study is gathered. This information mainly refers to two main categories:

- General aspects of the city, such as the size, transport infrastructures, and connections. High-level information about transport stakeholders is also provided.
- Objectives defined by the case study leader as strategic improvements of the transport system. These objectives include, for example, transport optimization, enhancement of sustainable transport, cooperative incident management, reduction of congestion, actor cooperation, and others. The list of the objectives for each case study is reported in the sheet titled "NAME_GENERAL_INFO" of the related catalogue, along with a brief description.

For each case study, the data sources may be numerous and very different. They may provide, for example, data related to public transport timetables or traffic conditions. Others may be related to sensors of various types, infrastructure, transportation and traffic control systems, telecommunication systems, maps, and environmental data. The goal is to bring out aspects peculiar to the available data

sources for each means of transportation. For this reason, the second objective of Phase 1 consists in structuring the data catalogue to facilitate the description of all the available data sources.

The categorization defined in the context of the H2020 project Momentum³ has been adopted to guide the collection and description of the data sources. Table 4 reports the considered data source categories (i.e., Transport Supply, Transport Demand, Maps & Cartography, and Driving Behaviour). Sub-categories are defined to support a more specific categorization of the data sources. Finally, examples of data source types are given for each sub-category. These examples do not aim to provide a comprehensive list of data source types but only to guide their categorization.

Table 4 is reported in the excel sheet titled "CATALOGUE_GEN_INFO" of each TANGENT Catalogue.

Categories	Sub-categories	Data source type (Examples)
Transport supply	Public Transport Schedules and Lines	Public transport supply Fleet size and composition GPS traces from public transport vehicles Real-time public transport data
	Transport Network Geometry	Transport Network Databases, Open Street Maps, etc.
	Transport Network Performance	Traffic state and behaviour: volume, speed and occupancy data from loop detector, on-ramps, off-ramps, cameras, C-ITS infrastructure, GPS/FCD data, etc. Travel times: Google Maps API or similar services, travel times from Bluetooth sensors and/or GPS traces, travel times from ANPR, C-ITS infrastructure, etc.
	Traffic Control	Intersection control: type of intersection (stop sign, traffic signal, etc.), type of traffic signal, signal timing plan, etc. Ramp metering: type of ramp metering, metering plan, location of traffic sensors, etc.
	C-ITS Infrastructure	RSU location and technology, LTE location and technology, C-ITS available services, etc.
	On-demand Mobility	Taxi, Ride hailing and Demand Responsive Transport: stop, fleet, trips, etc..
	Shared Mobility	Station-based (bike/moto/scooter/car/...)-sharing system: stations locations and capacity, fleet size and composition, vehicle availability, etc. Free-floating (bike/moto/scooter/car/...)-sharing system: fleet size and composition, trips, GPS traces, operational zone, etc.
	Parking	Off-street parking lots On-street parking lots
Transport Demand	Public Transport	Aggregated: OD Matrix, Passenger counts from sensorized buses, etc.

³ <https://h2020-momentum.eu/>

Categories	Sub-categories	Data source type (Examples)
		Disaggregated: Smart Card Validations for urban buses, metro/underground, suburban trains, tram, etc.
	Shared Mobility	Aggregated: OD Matrix Disaggregated: (bike/moto/scooter/car/...)-sharing trips (origin, destination, timestamp, etc.), GPS traces, etc.
	On-demand Mobility	Aggregated: OD Matrix, use statistics, etc. from taxis, ride-hailing/sharing, demand responsive transport or other on-demand modes Disaggregated: FCD or trip data from taxis, ride-hailing/sharing, demand responsive transport or other on-demand modes
	Cycling	Aggregated: OD Matrix, Counts from sensors, Wi-Fi detectors Disaggregated: GPS tracking, Bluetooth detectors
	Pedestrian	Aggregated: OD Matrix, Counts from sensors, Wi-Fi detectors Disaggregated: GPS tracking, Bluetooth detectors
	Travel Mobility	Aggregated: OD Matrix or Aggregated Statistics from Household surveys, stated preference survey Disaggregated: Microdata from household surveys, revealed preferences (real trajectories) from Google Maps Timeline, trip chains (real trajectories) from smartphone crowd-sensed data (proprietary data)
	Car Traffic	Aggregated: OD Matrix, counts or volumes from loop detectors/sensors, link counts, intersection turning counts. Disaggregated: FCD (e.g., INRIX – proprietary data)
	Opportunistic Data Sources	Telecommunication Data (Call Detail Records), Social Media, etc.
	Parking	Aggregated: Periodical snapshots of parking occupancy. Disaggregated: parking transactions, vehicle entry/exit events
Maps & Cartography	Weather & Environmental data	Weather sensors, Weather Information services, Air Quality sensors, pollution data, etc.
	Transport related events	Roadworks, traffic incidents, etc.
	Social, Cultural or Sportive Events	Local events database
	Points of Interest	GIS database, Google Maps and similar APIs
Driving Behaviour	Naturalistic driving data (accelerations, braking, speeding, harsh events, etc.)	Data from telematics applications (e.g., O7 Telematics – proprietary data)

Table 4: List of potential data sources to be collected in the TANGENT project.

The available data sources for each case study must be uniquely identified in the data catalogue. A guide for defining such a unique identifier is also provided in "CATALOGUE_GEN_INFO". The convention for the data source's IDENTIFIER naming is as follows:

- Three capital letters to identify the case study the data source refers to;
 - “REN” is used for Rennes
 - “LIS” is used for Lisbon
 - “ATH” is used for Athens
 - “MAN” is used for the Manchester
- Four capital letters to identify the subject (e.g., transportation mode) the data source is about;
 - “BUST” means bus transportation mode
 - “METR” means metro transportation mode
 - “CARS” means car transportation mode
 - “TRAM” means tram transportation mode
 - “RAIL” means railway transportation mode
 - “BIKE” means bike-sharing transportation mode
 - “AIRQ” means air quality data
 - “DEMR” means demand-responsive transportation mode
 - “CITS” means cooperative intelligent transport systems
 - “TRAF” means road traffic data
 - “OTHD” means other data
- Two digits to indicate a progressive number of the data source.

For example, the identifier REN_BUST_03 indicated the third data source collected by TANGENT related to the bus transportation of Rennes.

3.2.2 Phase 2: TANGENT metadata profile

From a general point of view, metadata represents information about data. They enrich data by making it easy to define, use, and manage the information that can be extracted from them.

Metadata can be classified differently depending on the information they add. For example, some of them provide information about the sources of the data, their type and access rights; others provide information about how data is organized, its structure and relationships. Again, others provide more descriptive information related to the content represented by the data and the context they refer to.

The main advantage arising from the usage of metadata is that it allows users to obtain information that is both clear and consistent with the relevant context, about the provenance of the data, and it is also possible to track the data along with their sources.

Having settled down the role and relevance of metadata, the objective of Phase 2 consists in defining a reference metadata schema to describe all data homogeneously and uniformly, regardless of their source.

A state-of-the-art analysis has been performed to guide the definition of a reference metadata schema for the TANGENT catalogues. In particular, the analysis focused on DCAT, DCAT-AP, TransportDCAT-AP, and the Coordinated Metadata Catalogue.

- *DCAT*⁴ is an RDF vocabulary designed to describe data catalogues using a standardized set of classes and properties. It facilitates the interoperability among data catalogues published on the web and allows to easily find and use data. Using DCAT, publishers increase discoverability and applications easily consume metadata from multiple catalogues. The new updated DCAT

⁴ DCAT Version 2, W3C Recommendation (2020), <https://www.w3.org/TR/vocab-dcat-2/>

version 2 (DCAT-v2) adds to previous versions a data service cataloguing assistance or API and provides support in expressing relationships between data sources.

- *DCAT Application Profile (DCAT-AP)* [1] is an extension of DCAT developed to include classes and properties that the initial vocabulary did not consider. DCAT-AP is an ontology that contains mandatory, recommended, and optional classes and properties to make it easier to adapt to different domains.
- *TransportDCAT-AP* [2] extends DCAT-AP to improve the representation of metadata relevant for the transportation domain. The profile considers that the nature of the public transport data is geospatial, thus defining the metadata related to this type of information as mandatory. In addition, a specific set of admissible keywords is defined to standardise the range of metadata properties (i.e., the values that can be associated with property) and enable domain-related queries, e.g., considering specific transportation modes.
- *Coordinated Metadata Catalogue* [3]: National Access Points (NAPs) are established at the legislative level, according to EU Regulations 2017/1926 [4], 885/2013 [5], 886/2013 [6], and 2015/962, to make national transport data accessible. A specific effort of EU EIP WG NAP [7] is the harmonization of metadata of European NAPs. A first agreement is found in the form of the “Coordinated Metadata Catalogue”. It allows the description and discovery of the most relevant features of NAP datasets. It is worth noting that the “Coordinated Metadata Catalogue” already contains a mapping with standard vocabularies such as EU-EIP, DCAT-AP [8] and INSPIRE [9].

In compliance with ongoing initiatives and projects (e.g., the NAPCORE project⁵), DCAT-AP version 2.0.0 has been selected as the starting point to define the TANGENT metadata profile. An extension of DCAT-AP will be produced to cover the specificity of TANGENT data sources. The definition of the TANGENT metadata profile is an iterative work that will be finalized and documented in Deliverable 2.2 “Data-sharing governance model”. Table 5 reports the current version of the metadata profile adopted to describe data sources in the data catalogues. The metadata needed to fill the profile will be provided by the TANGENT case study leaders, data owner/producer and/or other stakeholders and complemented with context (location, date, time) and available information.

METADATA	DESCRIPTION
IDENTIFIER	It corresponds to the data source unique identifier. It is defined according to the guideline reported in the new sheet "CATALOGUE_GEN_INFO".
CATEGORIES	The field is defined by choosing an item from the provided list in the sheet "CATALOGUE_GEN_INFO".
SUB-CATEGORIES	The field is defined by choosing an item from the provided list in the sheet "CATALOGUE_GEN_INFO".
DATA SOURCE TYPE	It corresponds to the type of data source. Examples of such field are reported in the sheet "CATALOGUE_GEN_INFO".
DATA SOURCE NAME	It corresponds to the name of the data source. This name has to be as explicative as possible.
DATA SOURCE DESCRIPTION	It is a complete description of the data source content.
PROVIDER	It corresponds to the provider(s) of the data source (e.g., transport authorities, operator, etc) and/or the platform (e.g., NAP, open data platform) that stores/captures the data.

⁵ <http://www.napcore.eu/>

TYPE	It corresponds to the type of the data (e.g., static, historical, real-time).
TEMPORAL GRANULARITY	It is the temporal aggregation of the data (e.g., minutely, hourly, weekly, monthly or event-based).
TEMPORAL SCOPE	It is the timeframe of the data source (e.g., last year, 31st December 2020, 2017-2021, specific period, etc...).
GEOGRAPHICAL GRANULARITY	It represents the geographical aggregation of the data (e.g., streets, stops, neighbourhoods, lanes, city, region).
GEOGRAPHICAL SCOPE	It corresponds to the geographical coverage of the data (e.g., group of streets, specific area, city, metropolitan area, region, national).
AVAILABILITY	The field corresponds to the availability of data (e.g., open data, available only under agreement, released with specific licenses).
ACCESSIBILITY	It specifies how to access the data source (e.g., direct download, API).
DATA SOURCE URL	It corresponds to the URL of the data source.
DATA SCHEMA	It is the schema of the data source (e.g., GTFS, NeTEx, TransXchange, SIRI, etc) or, if a custom one is used, it corresponds to the link to the schema description document (e.g., swagger).
FORMAT	It refers to the format of the data (e.g., XML, CSV, JSON).

Table 5: The TANGENT metadata profile

3.2.3 Phase 3: Data collection

For each TANGENT case study, the data collection process is carried out by involving the case study leaders in describing the public and private data sources in the data catalogue according to the defined metadata profile (Phase 2). To facilitate the activity, each catalogue has been structured in different sheets, each dedicated to a specific transportation mode. Each sheet contains only the metadata description of data sources related to that transportation mode. For example, Table 6 shows the excel sheet related to the bus data available for the Rennes case study.

METADATA	BUS DATA SOURCES - RENNES CASE STUDY		
IDENTIFIER	REN_BUST_01	REN_BUST_02	REN_BUST_03
DATA CATEGORY	TRANSPORT_SUPPLY	TRANSPORT_SUPPLY	TRANSPORT_SUPPLY
DATA SUB-CATEGORY	Transport Network Geometry	Public Transport Schedules and Lines	Public Transport Schedules and Lines
DATA SOURCE TYPE	Map	Lines of the buses	Real-time public transport data
DATA SOURCE NAME	topologie-des-points-darret-de-bus-du-reseau-star	parcours-des-lignes-de-bus-du-reseau-star	tco-bus-vehicules-position-tr
DATA SOURCE DESCRIPTION	Station topological map. List of bus stops in the STAR network, including their name,	Bus routes of the STAR network	Position of buses in circulation on the STAR network in real time

METADATA	BUS DATA SOURCES - RENNES CASE STUDY		
	municipality and geolocation		
PROVIDER	Rennes Metropole	Rennes Metropole	Star Data Explore
TYPE	Static/mapping	Static/mapping	Real time (every minute)
TEMPORAL GRANULARITY	NA	NA	minutely
TEMPORAL SCOPE	since 2015	since 2015	since 2014
GEOGRAPHICAL GRANULARITY	stop coordinates	detailed geometry of the route between stops	bus coordinates
GEOGRAPHICAL SCOPE	metropolitan	metropolitan	metropolitan
AVAILABILITY	open data	open data	open data
ACCESSIBILITY	direct download, API	direct download, API	direct download, API
DATA SOURCE URL	Topologie des points d'arrêt de bus du réseau STAR — Open Data Rennes (rennesmetropole.fr)	Parcours des lignes de bus du réseau STAR — Open Data Rennes (rennesmetropole.fr)	Position des bus en circulation sur le réseau STAR en temps réel — STAR Data Explore
DATA SCHEMA	Not available	Not available	Not available
FORMAT	csv, json, excel, geo-json, shapefile, kml	csv, json, excel, geo-json, shapefile, kml	csv, json, excel, geo-json, shapefile, kml

Table 6: Example of Bus data sheet of Rennes Data Catalogue

The data collection is a process that requires continuous interactions with case study leaders, who contribute with information and data, as well with as reviews and updates. In this context, the TANGENT metadata profile is also subject to change since new metadata could become relevant for a better description of the available data sources. For this reason, iterations between Phase 2 and Phase 3 have been performed.

The TANGENT catalogues must be considered as the first consolidated release of Phase 3, i.e., the first map of available data sources for each TANGENT case study. However, during the following stages of the project, additional data requirements may arise, requiring further data collection and the update of the data catalogues.

3.2.4 Phase 4: TANGENT catalogue analysis

The TANGENT catalogues are analyzed in Phase 4. The analysis aims to verify the completeness, accuracy, and consistency of each metadata description included in the catalogues. Particular attention is given to the validation of the data source URL to control the data accessibility. Collaboration with the

case study leaders has been requested to enhance the metadata quality and make the catalogues as complete as possible. Annexes I-IV report the main information collected through the TANGENT data catalogues at the moment of writing this deliverable.

After this step of metadata quality checking, Phase 4 proceeds with the mapping between data requirements (see Figure 4) and the data sources included in the catalogues.

Since feedback in terms of system and user requirements from the stakeholders involved in the co-creation processes was necessary to define the data requirements, they have not been able to guide the data collection process from the beginning. To establish contingency measures and speed up the data collection process, the aforementioned data catalogues were established. Therefore, the data requirements have been used to filter the necessary data sources from the data catalogues and to identify gaps.

The rest of this section describes the current status of the mapping between the data requirements and available data sources for each case study. The mapping is shown through a dedicated table derived from Table 3 and structured as follows:

- Data requirements (column “**Data Requirement**”) clustered by topic (column “**Topic**”) relevant for the case study;
- Refinement information (column “**Refinement**”) specific for the case study;
- Type of data needed (column “**Type**”);
- The ID and a short description of the selected data sources to fulfil the data requirements (column “**Data Collection Mappings**”). The detailed metadata description of the data sources is available in the Annexes I-IV.

The mapping process is complemented by an additional activity performed with the case study leaders to identify national/regional data portals and initiatives that already collect mobility data and could be used to enhance the data collection in the following stages of the TANGENT project. Furthermore, scouting of data services available online (free and/or requiring an access fee) has been performed to identify a list of data sources that could fulfill specific data requirements.

The list of identified data portals and online data services is reported in Table 7. The possibility of using these data sources to cover data requirements for a case study is mentioned in the next sub-sections.

CASE STUDY	DATA PORTAL	DESCRIPTION
RENNES	FRENCH NAP ⁶	National Access Point for transport data related to Delegated Regulation 2017/1926
	TRANSPORT OBSERVATORY ⁷	The Authority's Observatory collects data from players in the rail transport, road passenger transport and motorway sectors (railway, coach, bus station and road development, highways). It publishes quarterly and annual activity reports as well as specific studies and surveys on these markets.

⁶ <https://transport.data.gouv.fr>

⁷ <https://www.autorite-transport.fr/observatoire-des-transport/jeux-de-donnees-en-open-data/>

CASE STUDY	DATA PORTAL	DESCRIPTION
	RUDI ⁸	The RUDI project is a European project (2019-2022) supported by Rennes Metropole and funded by the European Union as part of the UIA program⁹. RUDI's objective is to enable administrations, private companies, associations and researchers in Rennes Metropole to easily access a wide variety of data in order to produce services of general interest that respect privacy. The RUDI interface consists of a web portal for accessing data from the territory to promote their use and its functioning is based on the creation of a meta-catalogue of data. In the ambit of the TANGENT project, RUDI could act as the Rennes Metropole data provider of reference, i.e., as a single access point to the Rennes available data sources. The current implementation of RUDI¹⁰ contains 383 data sources, of which 48 tagged with the Mobility and Transportation thematic.
LISBON	PORTUGUESE NAP ¹¹	Portuguese NAP related to Delegated Regulations 885/2013, 886/2013, 962/2015, 1926/2017. The Multimodal Information of Portuguese NAP is under development. A limited set of data sources are available (10) represented with heterogeneous data model (GTFS, NeTEx).
	LISBOA ABERTA ¹²	Lisbon Open Data Portal with Mobility information (in Portuguese language) and data sources. The portal refers to 43 mobility data sources. No classification is provided for the data sources.
	TENtec INTERACTIVE MAP VIEWER ¹³	Interactive Map Viewer about C-ITS Stations. The information is represented on a map.
MANCHESTER	GB DATA SOURCES ¹⁴	Website listing numerous UK Transport data sources
	TRANSPORT DATA GOV ¹⁵	National Data Portal: access point of possible data sources (Delegated Regulations: 1926/2017, 962/2015, 886/2013, 885/2013)

⁸ <https://rudi.datarennes.fr/>

⁹ <https://uia-initiative.eu/en/uia-cities/rennes-metropole>

¹⁰ <https://rudi.bzh/catalog>

¹¹ <https://nap-portugal.imt-ip.pt/nap/home>

¹² <https://lisboaaberta.cm-lisboa.pt>

¹³ <https://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/map/maps.html>

¹⁴ <https://bustimes.org/data>

¹⁵ Results applying the *Transport* filter on <https://data.gov.uk>

CASE STUDY	DATA PORTAL	DESCRIPTION
	NATIONAL PUBLIC TRANSPORT ACCESS NODES (NapTAN) ¹⁶	National public transport access points of possible data sources. It gives access to UK Transport Stops Data for different transport modes (buses, tram, metro, rail stations, etc.)
ALL	OVERPASS API ¹⁷	The Overpass API is a read-only API that can be used to retrieve selected portions of the Open Street Map data.
	WEATHER APIs	Available data services online offering access to weather forecast data, alerts, and historical data: OpenWeatherMap API ¹⁸ (weather forecast), Weatherbit API ¹⁹ (weather forecast and alerts), AccuWeather API ²⁰ (weather conditions), Visual Crossing API ²¹ (weather forecast and historical data), AerisWeather API ²² (weather alerts, conditions, forecasts).
	HOLIDAYS APIs	Available data services online offering access to holidays data: Abstract Holidays API ²³ , Timeanddate Holidays API ²⁴ .

Table 7: Available data portals and online data services to identify additional data sources

3.2.4.1 Data Collection: Lisbon case study

Table 8 reports the current status of the mapping between data requirements and available data sources for the Lisbon case study.

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Lisbon
TRANSPORT NETWORK	ROAD TRANSPORT NETWORK (roads, limited access zones, etc)			Covered by OVERPASS API
	ROAD EQUIPMENT POSITION	Traffic lights, Traffic loops, Cameras, Other sensors		LIS_CITS_01 (RSU locations - GPS coordinates) LIS_CARS_06 (Intersection control info) LIS_CARS_07 (Speed and lane control signs,

¹⁶ <https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan>

¹⁷ https://wiki.openstreetmap.org/wiki/Overpass_API

¹⁸ <https://openweathermap.org/api>

¹⁹ <https://www.weatherbit.io/api>

²⁰ <https://developer.accuweather.com/>

²¹ <https://www.visualcrossing.com/weather-api>

²² <https://www.aerisweather.com/support/docs/api/>

²³ <https://www.abstractapi.com/api/holidays-api>

²⁴ <https://dev.timeanddate.com/holidays/>

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Lisbon
				traffic counters and classifiers info)
	PUBLIC TRANSPORT NETWORK	Stops for Bus, Tram, Train, Metro		LIS_TRAM_03 (CARRIS Schedules, Lines and Tram locations) LIS_BUST_02 (CARRIS Schedules, Lines and Bus locations)
	ROAD TRANSPORT NETWORK EVENTS (planned)	Positions, start and duration of the event	Real time	
			Historical	
	PUBLIC TRANSPORT NETWORK EVENTS (planned)		Real time	
			Historical	
	ROAD TRANSPORT NETWORK INCIDENTS (unplanned)	Positions, start and duration of the event	Real time	LIS_CITS_03 (C-ITS Declared incidents) LIS_CITS_05 (Traffic incidents info sharing)
			Historical	LIS_CITS_03 (C-ITS Declared incidents) LIS_CITS_05 (Traffic incidents info sharing)
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)		Real time	
			Historical	
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)	Traffic counters, Tolling transaction	Real time	
			Historical	LIS_CARS_01 (Motorways traffic counters) LIS_CARS_04 (Motorways tolling transactions) LIS_CARS_05 (Urban roads traffic counters)
	FLOATING VEHICLE DATA (GPS, mobile, etc)	C-ITS data	Real time	LIS_CITS_02 (C-ITS CFD)
			Historical	LIS_CITS_02 (C-ITS CFD)
	ROAD TRAVEL TIMES (external services, statistics, etc)	Waze, TomTom, Google Maps, etc. for Journey time		LIS_TRAM_04 (CARRIS Tram network performance) LIS_BUST_04 (CARRIS Bus network performance)
PUBLIC TRANSPORT	SCHEDULES AND LINES	Bus, Tram, Train, Metro		LIS_TRAM_03 (CARRIS Schedules, Lines and Tram locations) LIS_BUST_02 (CARRIS

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Lisbon
				Schedules, Lines and Bus locations)
	TICKET VALIDATION DATA (e.g., smart card usage)			LIS_TRAM_01 (CARRIS Tram Ticket validations) LIS_BUST_01 (CARRIS Bus Ticket validations)
	FLOATING PT VEHICLE DATA (GPS, mobile, etc)	Bus and Tram	Real time	LIS_TRAM_03 (CARRIS Schedules, Lines and Tram locations) LIS_BUST_02 (CARRIS Schedules, Lines and Bus locations)
			Historical	
	DELAYS		Real time	
			Historical	
	OCCUPANCY	Bus and Tram	Real time	
			Historical	LIS_TRAM_02 (CARRIS Tram Passengers Counts) LIS_BUST_03 (CARRIS Bus Passengers Counts)
INFLUENCING FACTORS	WEATHER DATA (measurements, e.g., temperature, humidity, air quality, etc)		Real time	<i>Covered by WEATHER APIs</i>
			Historical	<i>Covered by WEATHER APIs</i>
	WEATHER EVENTS		Real time	LIS_CITS_04 (C-ITS Weather event warnings)
			Historical	LIS_CITS_04 (C-ITS Weather event warnings)
	FORECASTED WEATHER DATA		Real time	<i>Covered by WEATHER APIs</i>
			Historical	<i>Covered by WEATHER APIs</i>
	HOLIDAY/SEASONAL CALENDAR			<i>Covered by HOLIDAYS APIs</i>
	PLANNED EVENTS (sport, entertainment, etc)		Real time	
			Historical	
OTHERS	DATA ABOUT OTHER TRANSPORTATION MODES (shared mobility, taxis, etc)			

Table 8: Mapping between data requirements and available data sources (LISBON case study)

The current status of the data collection for the Lisbon case study can be summarized as follows:

- 18 data sources have been selected from the Lisbon data catalogue to cover the data requirements. All the selected data sources are currently not accessible. Further investigation and potential specific agreements with the data owners/providers are needed.
- The main gaps are represented by the lack of data sources related to:
 - real-time traffic measurements;
 - data related to events/incidents on the road transport and public transport network;
 - data related to public transport delays.
- The TANGENT project is going to solve the mentioned gaps and issues through the following actions:
 - continuous scouting of available data sources in collaboration with Lisbon's case study leader and stakeholders;
 - continuous working on making specific data sources available;
 - definition and signature of NDA for having access to private data from organizations not directly involved in the project;
 - creation of data sources for testing purposes (e.g., creation of historical data sources through the collection of real-time data, computation of delays from public transport floating vehicle data and expected schedule);
 - re-discussion of TANGENT services or specific functionalities to be tested in Lisbon.

3.2.4.2 Data Collection: Manchester case study

Table 9 reports the current status of the mapping between data requirements and available data sources for the Manchester case study.

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Manchester
TRANSPORT NETWORK	ROAD TRANSPORT NETWORK (roads, limited access zones, etc)			Covered by OVERPASS API
	ROAD EQUIPMENT POSITION	CCTV, Sensors, Traffic Signals, VMS		TFGM_CARS_02 (Mapped location of CCTV and VMS - Location of assets - Google my maps) TFGM_CARS_03 (Mapped location of traffic signal assets) TFGM_CARS_05 (Real time traffic flow and vehicle classification on Highway network, sensor location)
	PUBLIC TRANSPORT NETWORK	Bus and Tram		TFGM_BUST_02 (Fixed bus stop locations)
	ROAD TRANSPORT NETWORK EVENTS (planned)	Positions, start and duration of the event	Real time	TFGM_CARS_12 (Future planned roadworks, sporting events/concerts road closures, diversions)
			Historical	TFGM_CARS_12 (Future planned roadworks, sporting

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Manchester
				events/concerts road closures, diversions)
	PUBLIC TRANSPORT NETWORK EVENTS (planned)		Real time	<i>Covered by Road Transport Network Incidents for buses and trams</i>
			Historical	<i>Covered by Road Transport Network Incidents for buses and trams</i>
	ROAD TRANSPORT NETWORK INCIDENTS (unplanned)	Positions, start and duration of the event	Real time	TFGM_CARS_17 (Crowd sourced info from app users on incidents they log about the network) TFGM_OTH01_01 (Collision data each year)
			Historical	TFGM_CARS_13 (Incident location, duration, impact)
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)		Real time	<i>Covered by Road Transport Network Incidents for buses and trams</i>
			Historical	<i>Covered by Road Transport Network Incidents for buses and trams</i>
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)	Traffic counters	Real time	TFGM_CARS_05 (Real time traffic flow and vehicle classification on Highway network, sensor location) TFGM_CARS_06 (Highway count/classification, turning movement, speed, journey time, sensor location) TFGM_CARS_14 (Real time flow, speed, journey time for strategic road network - carriageway loops)
			Historical	
	FLOATING VEHICLE DATA (GPS, mobile, etc)		Real time	TFGM_CARS_04 (Real time journey time data from Bluetooth sensors)
			Historical	
	ROAD TRAVEL TIMES (external services, statistics, etc)	Journey time for routes		TFGM_CARS_14 (Real time flow, speed, journey time for strategic road network - carriageway loops) TFGM_CARS_16 (Crowd sourced journey time data

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Manchester
				obtained by Waze tracking its app users)
PUBLIC TRANSPORT	SCHEDULES AND LINES	Bus and Tram		TFGM_BUST_01 (Current bus route and scheduled bus timetable)
	TICKET VALIDATION DATA (e.g., smart card usage)			
	FLOATING PT VEHICLE DATA (GPS, mobile, etc)	Bus	Real time	
			Historical	
	DELAYS	Tram	Real time	TFGM_TRAM_01 (Real time departure data and network status)
			Historical	
	OCCUPANCY	Tram	Real time	
			Historical	
INFLUENCING FACTORS	WEATHER DATA (measurements, e.g. temperature, humidity, air quality, etc)		Real time	TFGM_AIRQ_01 (Sensors for hourly ambient air quality concentration) <i>Covered by WEATHER APIs</i>
			Historical	<i>Covered by WEATHER APIs</i>
	WEATHER EVENTS		Real time	<i>Covered by WEATHER APIs</i>
			Historical	<i>Covered by WEATHER APIs</i>
	FORECASTED WEATHER DATA		Real time	TFGM_OTHD_2 (Met Office Weather DataHub)
			Historical	<i>Covered by WEATHER APIs</i>
	HOLIDAY/SEASONAL CALENDAR			<i>Covered by HOLIDAYS APIs</i>
	PLANNED EVENTS (sport, entertainment, etc)		Real time	TFGM_CARS_12 (Future planned roadworks, sporting events/concerts road closures, diversions)
			Historical	TFGM_CARS_12 (Future planned roadworks, sporting events/concerts road closures, diversions)
OTHERS	DATA ABOUT OTHER TRANSPORTATION			TFGM_PEDE_01 (Pedestrian volumes and path) TFGM_BIKE_01 (Cycle trips,

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Manchester
	MODES (shared mobility, taxis, etc)			availability of bike hire, dock locations)

Table 9: Mapping between data requirements and available data sources (MANCHESTER case study)

The current status of the data collection for the Manchester case study can be summarized as follows:

- 18 data sources have been selected from the Manchester data catalogue to cover the data requirements. All the 18 selected data sources are open data accessible with the URL reported in the Manchester catalogue;
- The main gaps are represented by the lack of data sources related to public transport data (floating vehicle data, ticket validation, occupancy).
- The TANGENT project is going to solve the mentioned gaps and issues through the following actions:
 - continuous scouting of available data sources in collaboration with Manchester's case study leader and stakeholders;
 - definition and signature of NDA for having access to private data from organizations not directly involved in the project;
 - creation of data sources for testing purposes (e.g., creation of historical data sources through the collection of real-time data);
 - re-discussion of TANGENT services or specific functionalities to be tested in Manchester.

3.2.4.3 Data Collection: Rennes case study

Table 10 reports the current status of the mapping between data requirements and available data sources for the Rennes case study.

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Rennes
TRANSPORT NETWORK	ROAD TRANSPORT NETWORK (roads, limited access zones, etc)	Reserved bus lanes		REN_BIKE_06 (Cycling facilities and traffic calming zones) <i>Covered by OVERPASS API</i>
	ROAD EQUIPMENT POSITION	Traffic lights, Traffic loops		
	PUBLIC TRANSPORT NETWORK	Stops for Bus, Tram, Train, Metro		REN_BUST_01 (Station topological map. List of bus stops in the STAR network, including their name, municipality and geolocation) REN_BUST_06 (Geographical data of the STAR network: physical stops)
	ROAD TRANSPORT NETWORK EVENTS (planned)	Positions, start and duration of the event	Real time	REN_TRAF_03 (Winter traffic circulation)
			Historical	REN_TRAF_03 (Winter traffic circulation)

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Rennes
	PUBLIC TRANSPORT NETWORK EVENTS (planned)	Positions, start and duration of the event	Real time	
			Historical	
	ROAD TRANSPORT NETWORK INCIDENTS (unplanned)	Positions, start and duration of the event	Real time	REN_BUST_07 (Real-time traffic alerts on STAR network lines) REN_TRAF_01 (Winter traffic conditions)
			Historical	
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)	Positions, start and duration of the event	Real time	REN_BUST_07 (Real-time traffic alerts on STAR network lines) REN_METR_01 (Status of the STAR metro lines network in real time)
			Historical	
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)	Traffic flows, Congestion of Reserved Bus Lanes	Real time	REN_TRAF_02 (Gross Floating car data data on the intra-Metropolitan perimeter (Speed, Transit time) REN_TRAF_07 (Traffic data from counting stations in Traffic Management Centre control - Metropolitan) REN_TRAF_08 (Traffic count on national and county council roads - DIRO) REN_TRAF_09 (Traffic light plans, ramp access status)
			Historical	REN_TRAF_07 (Traffic data from counting stations in Traffic Management Centre control - Metropolitan) REN_TRAF_09 (Traffic count on national and departmental roads - DIRO)
	FLOATING VEHICLE DATA (GPS, mobile, etc)	Floating Car Data (FCD)	Real time	
			Historical	
	ROAD TRAVEL TIMES (external services, statistics, etc)	Floating Car Data (FCD)		
PUBLIC TRANSPORT	SCHEDULES AND LINES	Bus and Metro		REN_BUST_02 (Bus routes of the STAR network)

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Rennes
				REN_BUST_05 (List of services on the bus routes of the STAR network, including the order of the service) REN_METR_02 (Schedules and lines for metro)
	TICKET VALIDATION DATA (e.g., smart card usage)	Bus and Metro		REN_METR_03 (Ticketing data from the Cleunay Station)
	FLOATING PT VEHICLE DATA (GPS, mobile, etc)	Bus	Real time	REN_BUST_03 (Position of buses in circulation on the STAR network in real-time) REN_BUST_04 (Geolocation of the STAR network buses in real time that are in service with information on the new ticketing equipment)
			Historical	
	DELAYS	Bus and Metro	Real time	
			Historical	
	OCCUPANCY	Bus and Metro (ticketing data)	Real time	REN_METR_03 (Ticketing data from the Cleunay Station)
			Historical	REN_METR_03 (Ticketing data from the Cleunay Station) REN_BUST_08 (Maximum ridership level per STAR line) REN_BUST_09 (Monthly aggregated Bus and metro occupancy data)
INFLUENCING FACTORS	WEATHER DATA (measurements, e.g. temperature, humidity, air quality, etc)	Weather information and air quality data	Real time	REN_AIRQ_01 (Air quality from vehicles) REN_AIRQ_02 (Air quality from fix stations) REN_AIRQ_03 (Air quality from LoRa sensors) REN_OTHD_04 (Current conditions API)
			Historical	REN_AIRQ_01 (Air quality from vehicles) REN_AIRQ_02 (Air quality from fix stations) REN_AIRQ_03 (Air quality from LoRa sensors)

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Rennes
	WEATHER EVENTS		Real time	REN_OTHD_06 (Weather alarms API)
			Historical	Covered by WEATHER APIs
	FORECASTED WEATHER DATA		Real time	REN_OTHD_05 (Forecast API)
			Historical	Covered by WEATHER APIs
	HOLIDAY/SEASONAL CALENDAR	National education minister holiday calendar		REN_OTHD_02 (Le calendrier scolaire)
	PLANNED EVENTS (sport, entertainment, etc)	Specific events, SPACE Event, Stade Rennais matches	Real time	REN_OTHD_01 (Stade Rennais games season calendar) REN_OTHD_03 (SPACE event dates)
			Historical	REN_OTHD_01 (Stade Rennais games season calendar) REN_OTHD_03 (SPACE event dates)
OTHERS	DATA ABOUT OTHER TRANSPORTATION MODES (shared mobility, taxis, etc)			REN_BIKE_01 (Station locations) REN_BIKE_02 (Stations topology) REN_BIKE_03 (Real-time status of stations) REN_BIKE_04 (Counts of cycles) REN_BIKE_05 (Counts of pedestrian) REN_BIKE_07 (Secure bicycle parks) REN_BIKE_08 (Real time bike service occupation)

Table 10: Mapping between data requirements and available data sources (RENNES case study)

The current status of the data collection for the Rennes case study can be summarized as follows:

- 35 data sources have been selected from the Rennes data catalogue to cover the data requirements. Among the 35 selected data sources:
 - 28 data sources are open data accessible with the URL reported in the Rennes catalogue;
 - 7 data sources are currently not accessible and further investigation and potential specific agreements with the data owners/providers are needed.
- The main gaps are represented by the lack of data sources related to:
 - road equipment position;

- planned events (e.g., maintenance) related to road and public transport;
- additional floating vehicle data not related to public transport;
- data related to public transport delays.
- The TANGENT project is going to solve the mentioned gaps and issues through the following actions:
 - continuous scouting of available data sources in collaboration with Rennes' case study leader and stakeholders;
 - definition and signature of NDA for having access to private data from organizations not directly involved in the project;
 - creation of data sources for testing purposes (e.g., creation of historical data sources through the collection of real-time data, computation of delays from public transport floating vehicle data and expected schedule);
 - re-discussion of TANGENT services or specific functionalities to be tested in Rennes.

3.2.4.4 Data Collection: Athens case study

Table 11 reports the current status of the mapping between data requirements and available data sources for the Athens case study.

According to the DoA, TANGENT is planning a virtual case study for Athens. For this reason, some data sources will be generated for the specific purposes of the case study.

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Athens
TRANSPORT NETWORK	ROAD TRANSPORT NETWORK (roads, limited access zones, etc)			ATH_CARS_01 (Athens road network Open Street Map) ATH_CARS_02 (Athens road network NTUA)
	ROAD EQUIPMENT POSITION	Traffic loops		ATH_CARS_05 (Loop detectors positions)
	PUBLIC TRANSPORT NETWORK	Bus, Metro		ATH_BUST_02 (Bus station maps) ATH_METR_02 (Metro network)
	ROAD TRANSPORT NETWORK EVENTS (planned)			<i>To be generated for the virtual case study</i>
	ROAD TRANSPORT NETWORK INCIDENTS (unplanned)			<i>To be generated for the virtual case study</i>
	PUBLIC TRANSPORT NETWORK INCIDENTS (unplanned)			<i>To be generated for the virtual case study</i>
ROAD TRANSPORT	ROAD TRAFFIC MEASUREMENTS (traffic occupancy, speed, flow)		Real time	ATH_CARS_03 (Traffic data about flow, speed, loop loc)
			Historical	
PUBLIC TRANSPORT	SCHEDULES AND LINES	Bus, Metro		ATH_BUST_01 (Bus lines timetables) ATH_BUST_03 (Bus line routes)

Topic	Data Requirement	Refinement	Type	Data Collection Mappings Athens
				ATH_METR_01 (Metro timetables)
	TICKET VALIDATION DATA (e.g., smart card usage)			ATH_BUST_04 (Bus demand, passengers per stop) ATH_METR_03 (Metro demand, passengers per station)
INFLUENCING FACTORS	PLANNED EVENTS (sport, entertainment, etc)			To be generated for the virtual case study
OTHERS	AREA OF DCP POLICY ZONE			To be generated for the virtual case study
	DRT FLEET CHARACTERISTICS			To be generated for the virtual case study

Table 11: Mapping between data requirements and available data sources (ATHENS case study)

The current status of the data collection for the Athens case study can be summarized as follows:

- 11 data sources have been selected from the Athens data catalogue to cover the data requirements. Among the 11 selected data sources:
 - 8 data sources are open data accessible with the URL reported in the Athens data catalogue;
 - 3 data sources are currently not accessible and further investigation and potential specific agreements with the data owners/providers are needed.
- The selected data sources currently fulfill all the data requirements identified. Additional data requirements may emerge during the assessment of the Athens transport model, i.e., for the calibration of the model.
- 6 data sources will be generated by the TANGENT project for the specific scenarios of the Athens virtual case study.
- The TANGENT project is going to solve the mentioned gaps and issues through the following actions:
 - continuous scouting of available data sources in collaboration with Athens's case study leader and stakeholders;
 - definition and signature of NDA for having access to private data from organizations not directly involved in the project;
 - creation of data sources for testing purposes (e.g., generation of data sources for the virtual case study, creation of historical data sources through the collection of real-time data).

4 Data Interoperability

This chapter discusses the data interoperability issues that WP2 should address in the definition of a TANGENT solution for data harmonisation and fusion. We propose a structured analysis considering the five challenges of data interoperability depicted in Figure 5.



Figure 5: Challenges to implement data interoperability

Each challenge can be described using a question:

- LOCATE: which data is available and where?
- ACCESS: how to obtain the data needed?
- HARMONISE: how to convert data according to the needed data model?
- INTEGRATE: how to ensure different data sources can be merged?
- EXTRACT: how to consume harmonized and integrated data?

In the following sections, each challenge is analysed considering inputs from the literature [10][11][12][13] and the issues assessed within the TANGENT project during the collection of data sources made available for each case study.

The generic scenario analysed in this chapter is represented in Figure 6. Given a set of stakeholders, we want to enable interoperability of their data. Therefore, in this chapter, we define a data catalogue as a generic web portal that describes a set of data sources (datasets and data services) and possibly provides also hosting/storing capabilities. Considering different data catalogues (e.g., Data Catalogue 1 and 2), each of them may contain:

- datasets in different formats and/or using different data models (A, B, C)
- data services relying on different specifications and technologies (I, II)
- metadata describing data sources according to different profiles (d,e)

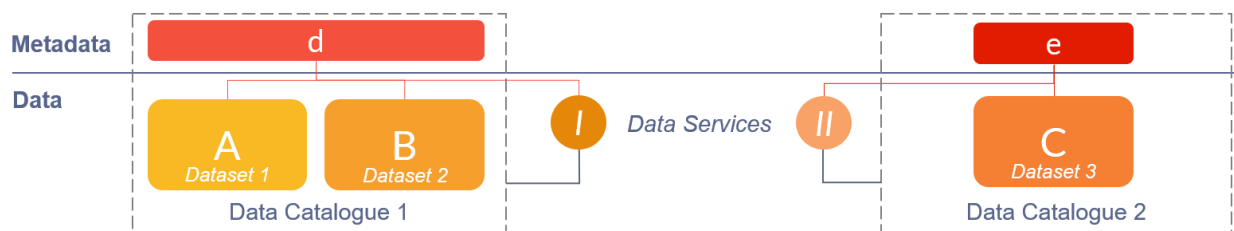


Figure 6: Generic scenario considering different data catalogues and data sources

4.1 Locate

The first challenge is the findability and discoverability of data. Data cannot be re-used and (made) interoperable if they cannot be found. For this reason, data catalogues/portals are implemented providing a solution to describe data sources, i.e., datasets and data services, through a set of metadata. The challenge is associated with the definition of a proper description of data sources and with several potential issues related to the interoperability of different data catalogues.

The findability of data can be implemented by having an online description of data through a set of common and machine-readable metadata. Potential issues are:

- [LOCATE-1] Data are not published and/or described online.
- [LOCATE-2] Data are published online without providing metadata.
- [LOCATE-3] Data are published using metadata that are not in a machine-readable format, i.e., they can not be automatically processed by automated agents (e.g., for querying).
- [LOCATE-4] Data are published with metadata information (e.g., metadata field value) provided in an unstructured format.
- [LOCATE-5] Data are published using custom metadata fields instead of relying on standardised vocabularies (may cause ambiguity issues).
- [LOCATE-6] The metadata quality is not sufficient, and metadata does not provide a proper description of the data resource.

The second aspect of this challenge is related to the interoperability of data catalogues and/or, in general, to solutions hosting data and their descriptions. Potential issues are:

- [LOCATE-7] Metadata cannot be programmatically retrieved and/or ingested from the data catalogue.
- [LOCATE-8] Metadata adopted by different catalogues rely on different vocabularies and require a harmonisation process to enable federated querying.
- [LOCATE-9] Metadata fields/properties adopting literal values (e.g., describing the category of a dataset) may cause string comparison issues (e.g., the same category called “public transport” in a catalogue and “public transportation” in another one).
- [LOCATE-10] Metadata may require manual inspection and/or editing, possibly related to incorrect usage of metadata fields/properties.

4.2 Access

Once data sources are located, the second challenge for enabling data interoperability is related to data accessibility. Data catalogues adopt different strategies for data access mainly associated with the

architectural choices for the hosting and storage of static and dynamic data sources. Potential issues related to this aspect are:

- [ACCESS-1] Data catalogues do not provide a direct mechanism (URL, API, etc.) to download/access the data resource.
- [ACCESS-2] Interface (URL, API, etc.) provided to download/access the data resource is no longer valid and/or accessible.
- [ACCESS-3] Access to the data requires different user authentication mechanisms (e.g., a token-based mechanism), possibly not directly handled by the data catalogue.
- [ACCESS-4] Access to the data requires specific authorizations for the user that may not be granted through a digital procedure (e.g., it requires signing an agreement between the parties).
- [ACCESS-5] Different versions of the same data resource are available, but it is not possible to access the specific version needed (e.g., URL pointing to a previous version of the data).
- [ACCESS-6] Access to the data is provided through an external API, possibly using a custom interface not documented in the data resource description.
- [ACCESS-7] Access to (real-time) data may require different types of interactions not supported by the system requesting the data (push/pull mechanism, asynchronous request, message queuing, database access and querying, etc.).
- [ACCESS-8] Access to (real-time) data may be granted only for a limited amount of time.

Further issues are related to the conditions for accessing and using data:

- [ACCESS-9] A policy regulating access to the data resource is not provided.
- [ACCESS-10] A set of conditions for the data resource usage is not provided.

4.3 Harmonise

The first two challenges analysed are related to the technical interoperability of systems that need to locate and access data using programmatic and automated procedures without intervention.

The third challenge is often the most difficult to address and comprises two objectives:

- Syntactic (structural) interoperability: compatibility of formats and/or protocols used to represent/exchange the data.
- Semantic interoperability: the ability of a system to understand (interpret) the information represented by the exchanged data unambiguously.

The potential issues are:

- [HARMONISE-1] Data are exchanged using different protocols (e.g., SOAP, RPC, REST).
- [HARMONISE-2] Data are represented using different formats (e.g., XML/CSV/JSON).
- [HARMONISE-3] Data are modelled by adopting different data models/standards.
- [HARMONISE-4] Data are modelled by adopting different versions of the same data model/standard.
- [HARMONISE-5] Data are modelled by adopting different conventions/(application) profiles for the same data model/standard.

As commented in Section 4.1, similar issues may also apply to metadata represented using different metadata profiles, thus requiring a harmonisation process. The discussion of issues related to the harmonisation of data should be complemented by considering the ones related to the implementation

of syntactic and semantic interoperability of different data sources. The common approach is to adopt an any-to-one centralized solution based on a reference conceptual model that improves scalability with respect to point-to-point solutions [14], as shown in Figure 7.

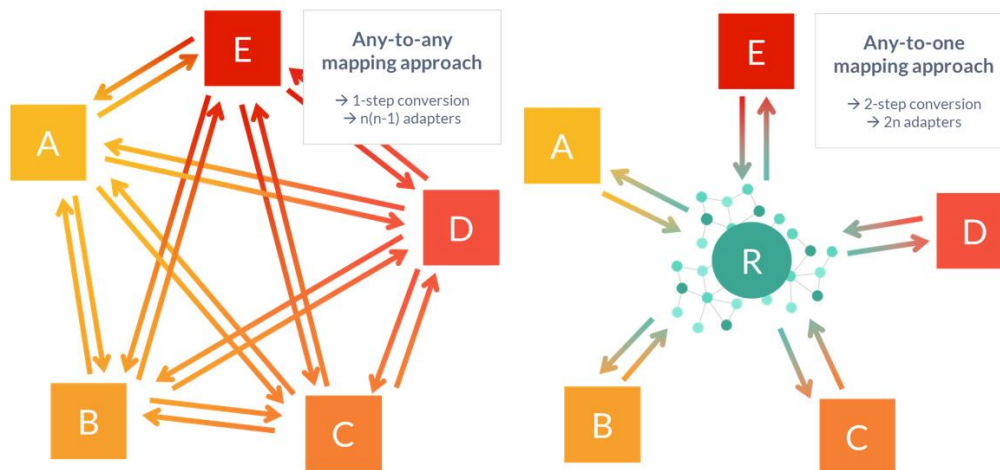


Figure 7: Any-to-any versus any-to-one mapping approaches

The potential issues related to the implementation of a harmonisation process are:

- [HARMONISE-6] The implementation of any-to-any interoperability leads to scalability issues due to an exponentially increasing number of mappings required with increasing data sources.
- [HARMONISE-7] A reference conceptual model, capable of representing the information exchanged/contained in different data sources, can be not already available.
- [HARMONISE-8] Different requirements on the format(s)/protocol(s) for the serialisation and sharing of the data harmonised through the conceptual model may exist.
- [HARMONISE-9] The mapping between each data resource and the identified conceptual model may not be trivial and require domain experts for its definition and assessment.
- [HARMONISE-10] In the definition of mappings, the reference conceptual model can be not able to represent all the information contained in the input data sources.

4.4 Integrate

The fourth challenge is related to data integration, which, after a harmonization process according to a reference conceptual model, may require further data processing to ensure their actual interconnection. Indeed, the harmonization process may cause the creation of “isolated sets of data” that are not actually interlinked and, therefore, effectively interoperable. Moreover, data may be represented using a shared semantic but without ensuring the possibility of having uniform processing.

Potential issues to be resolved for data integration are:

- [INTEGRATE-1] Different identifiers may be used to refer to the same entity and entity de-duplication may be needed.
- [INTEGRATE-2] Different literals (e.g., strings) representing the same value may exist and its processing may be needed (e.g., lowercase/uppercase, different abbreviations, different languages, etc.).

- [INTEGRATE-3] Different data sources may adopt different semantics to handle missing data (e.g., open-world assumption versus closed-world assumption, null versus missing value).
- [INTEGRATE-4] Data may be described considering different granularities (e.g., the average delay may refer to a specific train in a dataset or to the entire set of trains operated on a line in another dataset).
- [INTEGRATE-5] The same information may be contained in multiple data sources, and it may be necessary to discard some data to avoid inconsistencies (e.g., outdated information) in the integrated data.
- [INTEGRATE-6] Numerical data associated with the same variable may be represented considering different units of measurement.
- [INTEGRATE-7] Spatial data may be represented considering different Coordinate Reference Systems (CRS) and/or location referencing methods leading to different representations of the same location.
- [INTEGRATE-8] The assignment of a Primary Key (PK) strategy to a type/class of entities should guarantee that each data resource describing the entity has enough information to generate PKs of entities.
- [INTEGRATE-9] Data may be associated with contextual information (e.g., metadata) that is not directly represented but should be considered for their integration (e.g., the validity period of specific information).
- [INTEGRATE-10] Dynamic data may require specific integration strategies (e.g., on-demand/on-the-fly) with static data.

4.5 Extract

The fifth challenge is related to the consumption of harmonized and integrated data. The potential issues for this challenge refer to the implementation of a solution for the exploitation of the different data sources made interoperable. Potential issues are:

- [EXTRACT-1] It is required to implement access to specific subsets of the data.
- [EXTRACT-2] It is required to implement access to dynamic data as a stream.
- [EXTRACT-3] It is required to implement programmatic access to the data (e.g., REST API).
- [EXTRACT-4] It is required to implement access to the data through custom queries.
- [EXTRACT-5] It is required to implement access to different versions of the data.
- [EXTRACT-6] It is required to implement access to historical data (the implementation of a retention mechanism may be needed for dynamic data).
- [EXTRACT-7] The data should be exposed in a specific format (XML, JSON, etc.).
- [EXTRACT-8] The data should be exposed according to a specific data model/structure.
- [EXTRACT-9] Information about data provenance should be exposed together with the data.
- [EXTRACT-10] Specific agreements and/or technical solutions should be defined for the consumption of the data considering the access/usage policies of the data sources integrated.

4.6 TANGENT Data Interoperability Issues

This section reports a first analysis of the data interoperability issues to be addressed in TANGENT considering the data sources collected and presented in Section 3. In particular, we focus on the first two challenges, LOCATE and ACCESS, that cover potential issues associated with the data collection phase. Table 12 discusses relevant examples emerging from the data sources described in the TANGENT catalogues of the different case studies.

Identifier	Description	Example (e.g., dataset)
LOCATE-1	Data are not published and/or described online	The data resource LIS_CITS_01 is currently not published nor described online, thus requiring a manual process to request access and information about it.
LOCATE-2	Data are published online without providing metadata.	The data resource TFGM_OTHD_02 is associated with different datasets about Road Safety that are published online without detailed metadata information describing each file.
LOCATE-3	Data are published using metadata not in a machine-readable format, i.e., automated agents cannot automatically process them (e.g., for querying).	The data resource TFGM_TRAM_01 represents a data service accessible through API documented through web documentation and not described using a machine-readable format.
LOCATE-4	Data are published with metadata information (e.g., metadata field value) provided in an unstructured format	The value of the publisher and contact point metadata fields for the data resource REN_AIRQ_01 is provided as not structured text, i.e., parsing of contact information is necessary from a single literal value.
LOCATE-5	Data are published using custom metadata fields instead of relying on standardised vocabularies, which may cause ambiguity issues.	The data resource TFGM_BUST_01 is described according to a custom metadata vocabulary described online.
LOCATE-6	The metadata quality is not sufficient, and metadata does not provide a proper description of the data resource	The data resource ATH_METR_03 is described considering a limited set of metadata.
LOCATE-7	Metadata cannot be programmatically retrieved and/or ingested from the data catalogue.	The data resource ATH_BUST_03 is described using a set of metadata that are shown in the webpage, but cannot be programmatically retrieved (e.g., using an API).
LOCATE-8	Metadata adopted by different catalogues rely on different vocabularies and require a harmonisation process to enable federated querying.	Data sources for Athens are described in different data catalogues: catalog.hcapdata.gr (e.g., data source ATH_BUST_02) and data.gov.gr (e.g., data source ATH_BUST_04).
LOCATE-9	Metadata fields/properties adopting literal values (e.g., describing the category of a dataset) may cause string comparison issues (e.g., the same category called “public transport” in a catalogue and “public transportation” in another one).	The literal values for metadata fields associated with the data resource REN_BUST_01 are provided in French and may cause string comparison issues with values described in different languages.

Identifier	Description	Example (e.g., dataset)
LOCATE-10	Metadata may require manual inspection and/or editing, possibly related to incorrect usage of metadata fields/properties	The data resource REN_TRAF_02 has a metadata field reporting an update frequency of 5 minutes while the description reports a 3 minutes delay for the information to be updated, it should be checked whether the real-time information is refreshed every 3 or 5 minutes.
ACCESS-1	Data catalogues do not provide a direct mechanism (URL, API, etc.) to download/access the data resource.	The data resource ATH_METR_02 provides access to data through an online map but no direct mechanism to extract/export the data visualised is described.
ACCESS-2	The interface (URL, API, etc.) provided to download/access the data resource is no longer valid and/or accessible.	The data resource REN_METR_01 references a data access URL that is no longer valid.
ACCESS-3	Access to the data requires different user authentication mechanisms (e.g., a token-based mechanism), possibly not directly handled by the data catalogue.	The data resource ATH_BUST_04 requires a token-based mechanism to access the API.
ACCESS-4	Access to the data requires specific authorizations for the user that may not be granted through a digital procedure (e.g., it requires signing an agreement between the parties).	The data resource TFGM_TRAM_01 requires an agreement with the provider to get the credentials needed to access the data.
ACCESS-5	Different versions of the same data resource are available, but it is not possible to access the specific version needed (e.g., URL pointing to a previous version of the data).	The data resource TFGM_BUST_01 provides information on the updates made to bus timetables over time through a changelog, however, it is not possible to access a specific version (only the current version).
ACCESS-6	Access to the data is provided through an external API, possibly using a custom interface not documented in the data resource description.	The data resource REN_AIRQ_02 provides a web interface (UI) to select parameters and download the data needed but a description of the API interface for programmatic access is not provided.
ACCESS-7	Access to (real-time) data may require different types of interactions not supported by the system requesting the data (push/pull mechanism, asynchronous request, message queuing, database access and querying, etc.).	The provision of traffic-related events for Lisbon (data resource LIS_CITS_05) can be ensured only by implementing an endpoint accepting incoming requests (push) in a specific format.
ACCESS-8	Access to (real-time) data may be granted only for a limited amount of time.	The data resource REN_TRAF_02 provides access only to the latest values collected for traffic data. Previous values, e.g., to collect historical datasets, cannot be directly accessed

Identifier	Description	Example (e.g., dataset)
		and should be persisted by the client requesting it.
ACCESS-9	A policy regulating access to the data resource is not provided.	The policy to access the data sources LIS_CITS should be defined through a specific agreement.
ACCESS-10	A set of conditions for the data resource usage is not provided.	The conditions for the usage of the data sources LIS_CITS should be defined through a specific agreement.

Table 12: TANGENT data interoperability issues for the LOCATE and ACCESS challenges

The set of interoperability issues mentioned in Table 12 does not aim at being exhaustive but at providing practical examples of the first set of issues that should be handled by the WP2 TANGENT solution for data harmonisation and fusion.

After the finalisation of the data collection phase, the remaining three challenges, i.e., HARMONISE, INTEGRATE and EXTRACT, will be analysed in more detail considering the collected set of data sources. A preliminary example of issues considering the HARMONISE challenge can be provided considering the case of data sources collected for traffic-related information and represented using DATEX II²⁵:

- These data can be exchanged by different systems adopting different strategies only recently standardised through the DATEX II Exchange Specification²⁶ ([HARMONISE-1]).
- The DATEX II specification supports different formats for the exchange, such as XML and JSON ([HARMONISE-2]).
- The same information supported by the DATEX II specification can be represented using a different and/or custom schema/data model ([HARMONISE-3]).
- Different DATEX II data sources may expose data according to different versions of the DATEX II specification, e.g., v2.3 and v3.0 ([HARMONISE-4]).
- Different DATEX II data sources may implement different reference and/or national profiles²⁷ leading to the representation of the same information using different classes/properties and/or to a lack of information ([HARMONISE-5]).

The issues HARMONISE-6 to -10 should be considered in the definition of a TANGENT reference conceptual model and the associated harmonisation process.

The issues INTEGRATE-1 to -10 should be addressed considering different datasets having integration requirements.

Finally, the issues EXTRACT-1 to -10 will be considered according to the TANGENT architecture defined by WP6.

²⁵ www.datex2.eu

²⁶ <https://docs.datex2.eu/exchange/2020/rationale/index.html>

²⁷ https://www.datex2.eu/implementations/profile_directory

5 Conclusions

This deliverable has provided a first consolidated version of TANGENT data requirements extracted from the descriptions of the TANGENT services and case studies and from the specification of the components under development by the WP3-6 of the project. Nevertheless, during the following stages of the project, additional data requirements may arise from the refinement of the technical specification of the TANGENT platform, requiring further collection and validation of data requirements.

Moreover, this deliverable provides the current list of available data sources to fulfil the identified data requirements. The list has been defined through (i) the creation of data catalogues for the case studies with the involvement of the TANGENT case study leaders, (ii) the usage of the catalogues to identify the data sources that fulfil the data requirements, and (iii) refinement and validation of the mappings between data requirements and the data sources with the project partners.

The TANGENT catalogues (information currently collected available in Annexes I-IV) must be considered the first consolidated list of available data sources for each TANGENT case study. However, during the following stages of the project, additional data requirements may arise, requiring further data collection and the update of the data catalogues.

The analysis of the available data sources for each case study resulted in the identification of current gaps in fulfilling the data requirements and accessibility of data and licensing issues. The TANGENT project will face gaps and issues through the (i) continuous scouting of available data sources in collaboration with case study leaders and stakeholders and through the analysis of national/regional data portals and initiatives that already collect mobility data and could be used to enhance the data collection in the next stages of the TANGENT project, (ii) definition and signature of NDA for having access to private data from organizations not directly involved in the project, (iii) creation of data sources for testing purposes (e.g., creation of historical data sources through the collection of real-time data), and (iv) re-discussion of TANGENT services or specific functionalities to be tested in each demo site.

Finally, the deliverable has provided the complete set of potential data interoperability issues to be addressed by the TANGENT project through developing the WP2 solution for data harmonization and fusion (first release at M20). A first analysis of the available data sources resulted in the identification of data interoperability issues related to the data collection phase. These issues will be addressed as a priority in developing the WP2 solution to harmonise the process of locating and accessing data sources. In parallel, the retrieval and sharing of data samples (in raw format) from collected data sources will be performed to identify the set of interoperability issues to be addressed by the TANGENT project and to assess the quality and information content of each data source.

6 References

1. Van Nuffelen, B.: DCAT Application Profile for data portals in europe (DCAT-AP) v2.0.1. Tech. rep., SEMIC (2020), <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semantic/solution/dcat-application-profile-data-portals-europe/release/201-0>
2. Chaves, D., Rojas, J.A., Colpaert, P., Chamorro, J., Corcho, O.: A common metadata model for representing public transport at European level (2015), <https://github.com/cef-oasis/DCAT-AP/tree/master/TransportDCAT-AP>
3. EU EIP SA46. Coordinated Metadata Catalogue, https://www.its-platform.eu/wp-content/uploads/ITS-Platform/AchievementsDocuments/NAP/EU%20EIP_Coord.%20Metadata%20Catalogue_v2.0_191115.pdf
4. EU Delegated Regulation 2017/1926 https://eur-lex.europa.eu/eli/reg_del/2017/1926/oj
5. EU Delegated Regulation 2013/885 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013R0885&from=EN>
6. EU Delegated Regulation 2013/886 <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32013R0886>
7. EU EIP - Annual NAP Report 2020 <https://www.its-platform.eu/wp-content/uploads/ITS-Platform/AchievementsDocuments/NAP/EU%20EIP%20-%20National%20Access%20Points%20-%20annual%20report%202020.pdf>
8. DCAT-AP Vocabulary <https://www.w3.org/TR/vocab-dcat-2/#non-normative-namespaces>
9. INSPIRE Vocabulary <https://joinup.ec.europa.eu/collection/are3na/document/vocabulary-inspire-implementation-use>
10. J. Attard, F. Orlandi, S. Scerri, and S. Auer, “A systematic review of open government data initiatives,” *Government Information Quarterly*, vol. 32, no. 4, pp. 399–418, Oct. 2015, doi: 10.1016/j.giq.2015.07.006.
11. A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, “Challenges of data integration and interoperability in big data,” in *2014 IEEE International Conference on Big Data (Big Data)*, Oct. 2014, pp. 38–40. doi: 10.1109/BigData.2014.7004486.
12. F. Maali, R. Cyganiak, and V. Peristeras, “Enabling Interoperability of Government Data Catalogues,” in *Electronic Government*, Berlin, Heidelberg, 2010, pp. 339–350. doi: 10.1007/978-3-642-14799-9_29.
13. A. Carenini, A. Fiano, M. Scrocca, M. Comerio, and I. Celino, “Enabling Cross-Border Travel Offers Through National Access Point Federation via Metadata Harmonisation,” in *Proceedings of the 3rd International Workshop Semantics And The Web For Transport*, Online, September, Sep. 2021, vol. 2939. Accessed: Jun. 10, 2022. [Online]. Available: <http://ceur-ws.org/Vol-2939/#paper6>
14. M. Scrocca, M. Comerio, A. Carenini, and I. Celino, “Turning Transport Data to Comply with EU Standards While Enabling a Multimodal Transport Knowledge Graph,” in *The Semantic Web – ISWC 2020*, Cham, 2020, pp. 411–429. doi: 10.1007/978-3-030-62466-8_26.

Annex I

A snapshot summarising the information in the Athens catalogue is provided in Figure 8.

SUBJECT	IDENTIFIER	DATA CATEGORY	DATA SUB-CATEGORY	DATA SOURCE TYPE	DATA SOURCE NAME	DATA SOURCE DESCRIPTION	PROVIDER	TYPE	AVAILABILITY	ACCESSIBILITY
BUS	ATH_BUST_01	TRANSPORT_SUPPL Y	Public Transport Schedules and Lines	n/a	Lines timetables	bus lines timetables	OASA (authority-operator)	real time	TBD	API?
	ATH_BUST_02	TRANSPORT_SUPPL Y	Public Transport bus stops	n/a	Stations map	bus stations map	OASA (authority-operator)	historical	open	Download
	ATH_BUST_03	TRANSPORT_SUPPL Y	Public transport bus lines	n/a	Line routes	bus line routes	OASA (authority-operator)	historical	open	Download
	ATH_BUST_04	TRANSPORT_DEMAN D	Public Transport	Smart card	Passenger demand	bus demand (passengers per stop)	OASA (authority-operator)	real time	open	Download
METRO	ATH_METR_01	TRANSPORT_SUPPL Y	Public Transport Schedules and Lines	n/a	Timetables	metro timetables	STASY (authority-operator)	historical		TBD
	ATH_METR_02	TRANSPORT_SUPPL Y	Public Transport Schedules and Lines	n/a	Network	metro network	OASA (authority-operator)	static		TBD
	ATH_METR_03	TRANSPORT_DEMAN D	Public Transport	Smart card	Passengers demand	metro demand (passengers per station)	OASA (authority-operator)	real time	open	API
CAR	ATH_CARS_01	TRANSPORT_SUPPL Y	Transport Network Geometry	n/a	Road network	Athens road network	Openstreet Map	static	open	API
	ATH_CARS_02	TRANSPORT_SUPPK Y	Transport Network Geometry	Road transport network	Road network	Athens road network	NTUA	static	TBD	Offline
	ATH_CARS_03	TRANSPORT_DEMAN D	Car traffic	Loop detectors	Traffic data from detectors	traffic data (flow, speed, loop loc)	Region of Attica	real-time	open	API
	ATH_CARS_04	DRIVING_BEHAVIOR	Naturalistic driving data	Smartphone telematics application	Driving behavior data	driving behavior data	Oseven Telematics	historical	proprietary data	TBD
	ATH_CARS_05	TRANSPORT_DEMAN D	car traffic	Loop detectors position	Loop detectors position	Positions of Loop detectors	TBD	TBD	TBD	Offline
OTHER	ATH_OTH01_01	TRANSPORT_SUPPL Y	Traffic control	n/a	Traffic signal timeplan	traffic signal timeplan	Region of Attica	real-time	proprietary data	TBD

Figure 8: Snapshot of data sources described in the Athens catalogue

Annex II

A snapshot summarising the information in the Lisbon catalogue is provided in Figure 9.

SUBJECT	IDENTIFIER	DATA CATEGORY	DATA SUB-CATEGORY	DATA SOURCE TYPE	DATA SOURCE NAME	DATA SOURCE DESCRIPTION	PROVIDER	TYPE	AVAILABILITY	ACCESSIBILITY
BUS	US_BUST_01	TRANSPORT_DEMAND	Public Transport	Disaggregated: Ticket Validations for urban buses	CARRIS Bus Ticket validations	Bus ticket validation dataset from CARRIS (the public operator for urban trams and buses). The data will provide information on the #validations, per line, bus stop and direction.	CARRIS	historical	Available under agreement	
	US_BUST_02	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	Transport Supply & GPS traces from Bus vehicles	CARRIS Schedules, Lines and Bus locations	Bus lines, stop locations and real-time position of the Bus vehicles from CARRIS	CARRIS	static and real-time	open data	
	US_BUST_03	TRANSPORT_DEMAND	Public Transport	Aggregated: Occupancy data	CARRIS Bus Passengers Counts	Bus passenger counts from CARRIS. The data will provide on the #incoming / #outcoming passengers, per bus stop, line and direction - TBC	CARRIS	historical	Available under agreement	
	US_BUST_04	TRANSPORT_SUPPLY	Transport Network Performance	Traffic state and behavior	CARRIS Bus network performance	Bus network performance from CARRIS. The data will provide information on Commercial speed, Lost vehicle km, % scheduled trips operated, per bus line & direction for the desired timeframe - TBC	CARRIS	historical	Available under agreement	
TRAM	US_TRAM_01	TRANSPORT_DEMAND	Public Transport	Disaggregated: Ticket Validations for urban trams	CARRIS Tram Ticket validations	Tram ticket validation dataset from CARRIS (the public operator for urban Trams and Buses). The data will provide information on the #validations, per line, tram stop and direction.	CARRIS	historical	Available under agreement	
	US_TRAM_02	TRANSPORT_DEMAND	Public Transport	Aggregated: Occupancy data	CARRIS Tram Passengers Counts	Tram passenger counts from CARRIS. The data will provide on the #incoming / #outcoming passengers, per tram stop, line and direction. TBC	CARRIS	historical	Available under agreement	
	US_TRAM_03	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	Transport Supply & GPS traces from Tram vehicles	CARRIS Schedules, Lines and Tram locations	Tram lines, stop locations and real-time position of the Tram vehicles from CARRIS	CARRIS	static and real-time	open data	
	US_TRAM_04	TRANSPORT_SUPPLY	Transport Network Performance	Traffic state and behavior	CARRIS Tram network performance	Tram network performance from CARRIS. The data will provide information on Commercial speed, Lost vehicle km, % scheduled trips operated, per tram line & direction for the desired timeframe - TBC	CARRIS	historical	Available under agreement	
CAR	US_CARS_01	TRANSPORT_DEMAND	Car Traffic	Aggregated: Traffic counters	Motorways traffic counters	Traffic counters for A1, A5, A9, A12 motorways (from BCR), A8 (AEA) and 25 de Abril and Vasco da Gama bridges (Lusophone), at the Lisbon Metropolitan Area	Brisa Concessão Rodoviária (BCR), Autoestradas do Atlântico (AEA), Lusophone	historical data (few hours - 6 days delay, TBC)	available only under agreement	Ftp link
	US_CARS_02	TRANSPORT_DEMAND	Parking	Parking transactions	On Street Parking transactions	On Street Parking transactions	Via Verde	historical	available only under agreement	Ftp link
	US_CARS_03	TRANSPORT_DEMAND	Parking	Parking transactions	Off Street Parking transactions	Off Street Parking transactions	Via Verde	historical	available only under agreement	Ftp link
	US_CARS_04	TRANSPORT_DEMAND	Car Traffic	Disaggregated: Toll transactions	Motorways tolling transactions	Motorways tolling transactions for vehicles entering/exiting Lisbon through A1, A5, A9, A12 motorways (from BCR), A8 (AEA) and 25 de Abril and Vasco da Gama bridges (Lusophone)	Brisa Concessão Rodoviária, Autoestradas do Atlântico, Lusophone through Via Verde	historical (1-5 days delay), updated 1x/day	available only under agreement	Ftp link
	US_CARS_05	TRANSPORT_DEMAND	Car Traffic	Aggregated: Traffic counters	Urban roads traffic counters	Traffic counters for the urban roads inside the city of Lisbon	CMU/EMEL	historical	available only under agreement	
	US_CARS_06	TRANSPORT_SUPPLY	Traffic Control	Intersection control	Intersection control info	Intersection control information: type of intersection (stop sign, traffic signal, etc.), type of traffic signal, signal timing plan, etc.	CMU/EMEL	static data		
	US_CARS_07	TRANSPORT_SUPPLY	Traffic Control	Ramp metering	Speed and lane control signs, traffic counters and classifiers info	Ramp metering: type of ramp metering (speed and lane control), location of traffic sensors (counter, classifiers), etc.	CMU/EMEL	static data		
C-ITS	US_CITS_01	TRANSPORT_SUPPLY	C-ITS Infrastructure	RSU location	C-ITS RSU locations	RSU locations (GPS coordinates)	Private providers (Brisa, through A-to-Be)	Static	Open data	direct download
	US_CITS_02	TRANSPORT_DEMAND	Car Traffic	FCD from Connected vehicles	C-ITS CFD	Anonymous Car Floating Data from detected connected vehicles (position of the vehicles, instant speed, etc.), through the ITS-G5 infrastructure of RSUs connected directly to A-to-Be's Infrastructure (C-ITS network)	Private providers (Brisa, through A-to-Be)	historical data, real-time	Brisa/A-to-Be available under agreement	Cloud secure MQTT / Datex II node
	US_CITS_03	MAPS_CARTOGRAPHY	Transport related events	Declared incidents	C-ITS Declared incidents	Declared incidents from connected vehicles: vehicle breakdown, animal or person on the road congestions, etc.	Private providers (Brisa, through A-to-Be)	historical data, real-time	Brisa/A-to-Be available under agreement	Cloud secure MQTT / Datex II node
	US_CITS_04	MAPS_CARTOGRAPHY	Weather & Environmental Data	Weather information	C-ITS Weather event warnings	Weather event warnings published at A-to-Be's C-ITS network	Private providers (Brisa, through A-to-Be)	historical data, real-time	Brisa/A-to-Be available under agreement	Cloud secure MQTT / Datex II node
	US_CITS_05	MAPS_CARTOGRAPHY	Transport related events	Traffic incidents	Traffic incidents info sharing	Traffic incidents reported in real-time at the Traffic Management Centers of the concessions	Private providers (concessions) and the national road management entity through the National Access Point (NAP)	historical	TBD	TBD

Figure 9: Snapshot of data sources described in the Lisbon catalogue

Annex III

A snapshot summarising the information in the Manchester catalogue is provided in Figure 10.

SUBJECT	IDENTIFIER	DATA CATEGORY	DATA SUB-CATEGORY	DATA SOURCE TYPE	DATA SOURCE NAME	DATA SOURCE DESCRIPTION	PROVIDER	TYPE	AVAILABILITY	ACCESSIBILITY
BUS	TFGM_BUST_01	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	current bus route and scheduled bus timetable	Bus timetable	Scheduled bus timetable	Department for Transport	static	open	API
	TFGM_BUST_02	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	Fixed bus stop locations	Bus stops	Fixed bus stop locations	Department for Transport	static	open	API
	TFGM_BUST_03	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	real time bus location via AVL	Bus location	real time bus location	Department for Transport	real time	open	API
TRAM	TFGM_TRAM_01	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	Real time public transport data	Metrolink	real time departure data and network status	Metrolink - TFGM owned tram operator	real time	open data	API
BIKE	TFGM_BIKE_01	TRANSPORT_SUPPLY	Shared Mobility	Cycle trips	Beryl - cycle hire	Cycle trips, availability of bike hire, dock locations, via Beryl app who manage TFGM's cycle hire scheme				
PEDESTRIAN	TFGM_PEDE_01	TRANSPORT_SUPPLY	Pedestrian count data	Pedestrian volumes and path	Pedestrian counts	Videoanalytics sensor	Private provider https://vivacitylabs.com/	real time	Agreement with Vivacity - API	API
AIR QUALITY	TFGM_AIRQ_01	MAPS_CARTOGRAPHY	Weather & Environmental Data	Air quality sensors	Air quality	Sensors for hourly ambient air quality concentration	University of Manchester	real time	agreement	API
CAR	TFGM_CARS_01	TRANSPORT_SUPPLY	ITS Assets	Location of assets	Location of assets	Mapped location of Journey time sensors, Automatic traffic counters	TFGM	STATIC	Manchester TFGM Asset Map – Google My Maps	
	TFGM_CARS_02	TRANSPORT_SUPPLY	ITS Assets	Mapped location of CCTV and VMS	Location of assets	Location of assets	TFGM	STATIC	Assets – Google My Maps	
	TFGM_CARS_03	TRANSPORT_SUPPLY	Traffic signal locations mapped	Mapped location of traffic signal assets	Location of assets	location of assets	TFGM	STATIC	GMUTC signals maps (gmtu.gov.uk)	
	TFGM_CARS_04	TRANSPORT_SUPPLY	Traffic network performance	Bluetooth sensors	Journey time data	Real time journey time data from bluetooth sensors, sensor locations are mapped on the platform	Private provider https://www.drakewell.com/	real time	Agreement with Drakewell	API
	TFGM_CARS_05	TRANSPORT_SUPPLY	Traffic network performance	Fixed location automatic traffic counters - count and classification	Traffic information	Traffic flow and vehicle classification on Highway network includes mapped locations of sensors	Private provider https://www.drakewell.com/	real time	Agreement with Drakewell	API
	TFGM_CARS_06	TRANSPORT_SUPPLY	Traffic network performance	Video analytics sensors	Highway Information	Highway count/classification, turning movement, speed, journey time. Sensor location are mapped on the platform	Private provider https://vivacitylabs.com/	real time	Agreement with Vivacity	API
	TFGM_CARS_07	TRANSPORT_SUPPLY	Traffic network performance	Full turning count	Count/classification	count/classification	TFGM	historical	agreement with TFGM	
	TFGM_CARS_08	TRANSPORT_SUPPLY	Traffic flow survey	Origin destination survey	Vehicles origin/destination	origin destination of vehicles	TFGM	historical	agreement with TFGM	
	TFGM_CARS_09	TRANSPORT_SUPPLY	Traffic signal timing data	Traffic signal fault and SCOOT loop data	Urban traffic control (UTC) system	Urban traffic control (UTC) system	Private provider https://www.yunext.traffic.com/global/en/	real time	open	API
	TFGM_CARS_10	TRANSPORT_SUPPLY	Traffic signal timing data	Traffic signal phase and stage timing	Urban traffic control (UTC) system	Urban traffic control (UTC) system	Private provider https://www.yunext.traffic.com/global/en/	real time	under agreement	API
	TFGM_CARS_11	TRANSPORT_SUPPLY	Traffic signal timing data - TRL	Traffic signal phase and stage timing. GLOSA application	Traffic signal data	real time traffic signal data	Private provider https://trisoftware.com/	real time	open - need to request via TRL (Transport Research Laboratory)	
	TFGM_CARS_12	TRANSPORT_SUPPLY	Event data	TFGM - database, map layers	Future planned events	Future planned roadworks, sporting events, concerts road closures, diversions	TFGM	historical	open	API
	TFGM_CARS_13	TRANSPORT_SUPPLY	Incident data	TFGM - database, map layers	Incident location, duration, impact	Incident location, duration, impact	TFGM	historical	open	API
	TFGM_CARS_14	TRANSPORT_SUPPLY	Count/flow	Real time flow, speed, journey time for Strategic road network	Carriageway loops	Carriageway loops	National Highways https://nationalhighways.co.uk/	realtime	open	API
	TFGM_CARS_15	TRANSPORT_SUPPLY	Car park occupancy	Availability of spaces, status of car park	Detection system counting vehicles in/out	Detection system counting vehicles in/out	TFGM	real time	open	API
	TFGM_CARS_16	TRANSPORT_SUPPLY	Transport network performance	Real time journey time	Traffic data from Waze	Crowd sourced journey time data obtained by Waze tracking its app users	Waze for Cities	real time	Agreement with supplier reached	API
	TFGM_CARS_17	TRANSPORT_SUPPLY	Transport network performance	Incident alerts	Traffic data from Waze	Crowd sourced info from app users on incidents they log about the network	Waze for Cities	real time	Agreement with supplier reached	API
RAILWAY	TFGM_RAIL_01	TRANSPORT_SUPPLY	Public transport schedule	real time public transport	Service status/arrival times	real time service status/arrival times of train	Network Rail - national rail operator	real time		
OTHER	TFGM_OTH01	DRIVING_BEHAVIOR	Accidents Data	collision data - STATS19	Collision data	Collision Data each year in three tables "Collision, Casualty and Vehicle".	Department for Transport Met Office	Historical	OPEN	
	TFGM_OTH02	MAPS_CARTOGRAPHY	Weather & Environmental data	weather information services	Met Office Weather DataHub	Hourly, three hourly and daily weather forecast information given Latitude-Longitude coordinates	Met Office (national UK meteorological service)	Real-time/forecasted	OPEN (free tier or payment)	API

Figure 10: Snapshot of data sources described in the Manchester catalogue

Annex IV

A snapshot summarising the information for Rennes is provided in Figure 11 and Figure 12.

SUBJECT	IDENTIFIER	DATA CATEGORY	DATA SUB-CATEGORY	DATA SOURCE TYPE	DATA SOURCE NAME	DATA SOURCE DESCRIPTION	PROVIDER	TYPE	AVAILABILITY	ACCESSIBILITY
BUS	REN_BUST_01	TRANSPORT_SUPPL	Transport Network Geometry	Map	Topologie-des-points-darret-de-bus-du-reseau-star	Station topological map. List of bus stops in the STAR network, including their name, municipality and geolocation.	Rennes Metropole	Static/mapping	open data	direct download, API
	REN_BUST_02	TRANSPORT_SUPPL	Public Transport Schedules and Lines	Lines of the buses	Parcours-des-lignes-de-bus-du-reseau-star	Bus routes of the STAR network	Rennes Metropole	Static/mapping	open data	direct download, API
	REN_BUST_03	TRANSPORT_SUPPL	Public Transport Schedules and Lines	Real-time public transport data	Tco-bus-vehicules-position-tr	Position of buses in circulation on the STAR network in real time	STAR DATA EXPLORE	Real time (every minutes)	open data?	direct download, API or KEOLIS?
	REN_BUST_04	TRANSPORT_SUPPL	Public Transport Schedules and Lines	GPS traces from Public Transport vehicles	Tco-bus-vehicules-geoposition-suivi-new-billettique-tr	Geolocation of the STAR network buses in real time that are in service with information on the new ticketing equipment	STAR DATA EXPLORE	Real time (every minutes)	open data	direct download, API or KEOLIS?
	REN_BUST_05	TRANSPORT_SUPPL	Transport Network Performance	Bus services	Tco-bus-topologie-dessertes-td	List of services on the bus routes of the STAR network, including the order of the service.	Rennes Metropole	Static/mapping	open data	direct download, API
	REN_BUST_06	TRANSPORT_SUPPL	Transport Network Geometry	map	Equipement-accessibilite-arrets-bus	Geographical data of the STAR network : Physical stops	Rennes Metropole	Static/mapping	open data	direct download, API
	REN_BUST_07	TRANSPORT_SUPPL	Transport Network Performance	Traffic state and behavior	Alertes-traffic-en-temps-reel-sur-les-lignes-du-reseau-star	Real-time traffic alerts on STAR network lines	Rennes Metropole/KEOLIS	Real time (every minutes)	open data	direct download, API
	REN_BUST_08	TRANSPORT_SUPPL	Transport Network Performance	Occupancy data	Mkt-frequentation-niveau-freq-max-ligne	Maximum ridership level per STAR line	Rennes Metropole/KEOLIS	weekly updating	open data	direct download, API
	REN_BUST_09	TRANSPORT_SUPPL	Transport Network Performance	Occupancy data	Tco-billettique-frequentation-agregee-td	Monthly aggregated Bus and metro occupancy data	Rennes Metropole/KEOLIS	Monthly updating	Available under agreement?	direct download, FTP
BIKE	REN_BIKE_01	TRANSPORT_SUPPL	Shared-Mobility	Station location	Stations_vis	soft mobility data - localisation of STAR bike stations	Rennes Metropole	Static/mapping	Open Data	Direct download, API
	REN_BIKE_02	TRANSPORT_SUPPL	Transport Network Performance	Topology of stations	Topologie-des-stations-citiz	List of self-service bicycle stations in the STAR network, including their name, number of locations and geolocation.	Rennes Metropole	Static/mapping	Open Data	Direct download, API
	REN_BIKE_03	TRANSPORT_SUPPL	Shared-Mobility	Real time station status	Etat-des-stations-le-velo-star-en-temps-reel	Real-time status of self-service bicycle stations, including the number of bicycles available, the number of spaces available and their operational status. Please provide a data source type for this dataset.	Rennes Metropole	real time (every minutes)	Open Data	Direct download, API
	REN_BIKE_04	TRANSPORT_DEMAND	Cycling	Counts of cycles	Eco-counter-data	Data from meters installed by the City of Rennes on several sites. The counts include bicycles and pedestrians.	Rennes Metropole	Hours	Open Data	Direct download, API
	REN_BIKE_05	TRANSPORT_DEMAND	Pedestrian	Counts of pedestrian	Eco-counter-data	Data from meters installed by the City of Rennes on several sites. The counts include bicycles and pedestrians.	Rennes Metropole	Hours	Open Data	Direct download, API
	REN_BIKE_06	TRANSPORT_SUPPL	Shared-Mobility	Cycling facilities and traffic calming zones	Amenagements-velo-et-zones-de-circulation-apaisee	Existing state of specific facilities dedicated to cycles. Roads classified as traffic calming zones (30 km/h max).	Rennes Metropole	Static/mapping	Open Data	Direct download, API
	REN_BIKE_07	TRANSPORT_SUPPL	Shared-Mobility	Secure bicycle parks	Parcs-velos-securises-geres-par-rennes-metropole	Secure bicycle parks managed by Rennes Métropole	Rennes Metropole	Static/mapping	Open Data	Direct download, API
	REN_BIKE_08	TRANSPORT_SUPPL	Shared-Mobility	Real time bike service occupation	Bike service	Real time bike service occupation on the Rennes Metropol website: bicycle counting	Rennes Metropole	real time	data available under agreement	

Figure 11: Snapshot of data sources described in the Rennes catalogue (Part 1)

SUBJECT	IDENTIFIER	DATA CATEGORY	DATA SUB-CATEGORY	DATA SOURCE TYPE	DATA SOURCE NAME	DATA SOURCE DESCRIPTION	PROVIDER	TYPE	AVAILABILITY	ACCESSIBILITY
METRO	REN_METR_01	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	Real-time public transport data	Etat-des-lignes-de-metro-du-reseau-star-en-temps-reel	Status of the STAR metro lines network in real time	KEOLIS	Real time (every minutes)	Open data	direct download, API
	REN_METR_02	TRANSPORT_SUPPLY	Public Transport Schedules and Lines	Schedules and lines for metro and underground	Metro-du-reseau-star-traces-de-laxe-des-lignes	STAR metro lines	KEOLIS	Static/mapping	Open data	direct download, API
	REN_METR_03	TRANSPORT_DEMAND	Public Transport	Ticketing data	Ticketing data	Ticketing data from the Cleunay Station	KEOLIS	historical	TBD	CSV or FTP/API ?
AIR QUALITY	REN_AIRQ_01	OTHER	Air Quality	Air quality measurement	AQMO	Sensors (fix and mobile on vehicles)	Irisa	Real time	available under agreement	
	REN_AIRQ_02	OTHER	Air Quality	Air quality measurement	Air quality data from fix stations	Fix scientific instruments around Rennes Métropole	Airbreizh	Real time	open data	
	REN_AIRQ_03	OTHER	Air Quality	Air quality measurement	Air quality data from LoRa sensors	LoRa sensors	Rennes Métropole	Real time	available under agreement	API ?
TRAFFIC	REN_TRAF_01	DRIVING_BEHAVIOR	Traffic	Winter traffic conditions	Winter traffic conditions	Real time traffic conditions (good, bad, impossible, etc.) during winter (November 16 to March 8)	DIRO	real time (15 minutes)	open data	Direct download, API
	REN_TRAF_02	DRIVING_BEHAVIOR	Traffic	Traffic data	Real time traffic data	Gross FCD data on the intra-Metropolitan perimeter (V, TdP)	Autoroute trafic	real time (every 3 minutes)	open data	Direct download, API
	REN_TRAF_03	TRANSPORT_SUPPLY	Traffic	Winter traffic conditions	Winter traffic circulation	Winter traffic circulation (viabilité hivernale)	Rennes Métropole	real time (15 min)	open data	Direct download, API
	REN_TRAF_04	DRIVING_BEHAVIOR	Traffic	Traffic equipement	RM Traffic control equipement data	Supervision data from metropolitan traffic control equipment (controllers, traffic lights, sensors at traffic lights, etc.)	Rennes Métropole (DVGTS)	real time (1min)	available only under agreement	CSV (manual transmission)
	REN_TRAF_05	DRIVING_BEHAVIOR	Traffic	Traffic lights	Traffic lights plan data	Traffic light plans to be applied	Rennes Métropole (DVGTS)	real time (1min)	available only under agreement	CSV (manual transmission)
	REN_TRAF_06	DRIVING_BEHAVIOR	Traffic	Traffic lights	Traffic lights plan data	Applied traffic light plans, Top	Rennes Métropole (DVGTS)	real time (1min)	available only under agreement	CSV (manual transmission)
	REN_TRAF_07	DRIVING_BEHAVIOR	Traffic	Traffic station count	Comptages-routiers-station	traffic data from counting stations in PC control	Rennes Métropole (DVGTS)	1/5min	Open Data	Direct download, API
	REN_TRAF_08	DRIVING_BEHAVIOR	Traffic	Traffic count	DIRO traffic count data (Q, To, V)	Traffic count on national and departemental roads	DIRO (SGAT)	real time (1/5 min)	available only under agreement	CSV (manual) or FTP ?
	REN_TRAF_09	DRIVING_BEHAVIOR	Traffic	Traffic count	DIRO traffic count data (Q, To, V)	Traffic count on national and departemental roads	DIRO (SGAT)	historical	available only under agreement	CSV (manual) or FTP ?
	REN_TRAF_10	DRIVING_BEHAVIOR	Traffic	Traffic management	Keolis bus approach curves	Consideration of approach curves in the modulation of traffic light cycles to give priority to PT vehicles	KEOLIS	Real time	available only under agreement	CSV (manual) or FTP ?
OTHER	REN_OTHD_01	OTHER	Social, Cultural or Sportive Events	Stade Rennais's games	Stade rennais games season calendar	calendar of the ligue 1 season of Stade Rennais FC	LFP (French national football league)	static	open (data)	direct download from the ligue 1 website
	REN_OTHD_02	OTHER	Holiday/Seasonal Calendar	Holiday Calendar	Le calendrier scolaire	French national education Calendar (zone B for Rennes)	data.gouv French ministry of national education	static	open data	direct download from the datagouv website
	REN_OTHD_03	OTHER	Social, Cultural or Sportive Events	SPACE event	Date of the SPACE event	date of the SPACE event (international livestock show)	SPACE	static	open data	direct download from the SPACE Site
	REN_OTHD_04	OTHER	Weather	Current Weather	Current Conditions API	Current weather data for Rennes	Accuweather	real time	TBD	API
	REN_OTHD_05	OTHER	Weather	Weather Forecast	Forecast API	weather forecast for Rennes	Accuweather	real time	TBD	API
	REN_OTHD_06	OTHER	Weather	Weather Event	Weather Alarms API	Weather alerts for specific weather events	Accuweather	real time	TBD	API

Figure 12: Snapshot of data sources described in the Rennes catalogue (Part 2)