

D9.2

Data Management Plan. First release.



Deliverable administrative information

Deliverable number	D9.2
Deliverable title	Data Management Plan. First release.
Dissemination level	Public
Submission deadline	28/02/2022
Version number	2.0
Authors	Leire Serrano (Deusto) Marco Comerio (CEFRIEL) Antonia Azzini (CEFRIEL)
Internal reviewers	Véronique Rottier (ID4CAR) Athina Tympakianaki (Aimsun) Lara Moura (A-to-Be) Eleni Mantouka (NTUA) Celine Queron (Rennes Metropole) Siegfried Mercelis (IMEC) Emiliya Kamenova (POLIS)
Document approval	N/A

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_sTHj7fE4zXTToA

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

Executive summary

This document describes the Data Management Plan (DMP) for the TANGENT project. The DMP provides an analysis of the main elements of the data management policy that will be used throughout the project by the project partners, with regard to all the datasets that will be generated, harvested and/or used by the project.

In more detail this document covers the following:

1. TANGENT data identification and collection approach,
2. TANGENT overall dataset structure, including an overview of identified data sources and datasets
3. the TANGENT overall data management plan
4. dataset with personal data in TANGENT
5. Ethical aspects related to data management in TANGENT project.

As data management is an ongoing process along the duration of the TANGENT project and data management in the project is taking place in a dynamic environment, this document on hand is seen as a living document, which means that the document will be developed and maintained continuously over time.

Key words

Data Management Plan, personal data, artificial intelligence, data processing technologies, secure data sharing

Table of contents

DELIVERABLE ADMINISTRATIVE INFORMATION.....	1
EXECUTIVE SUMMARY	3
TABLE OF CONTENTS	4
1 LIST OF ABBREVIATIONS AND ACRONYMS	6
2 INTRODUCTION	7
3 DATASET STRUCTURE IN TANGENT	8
4 DATA MANAGEMENT IN TANGENT	9
5 DATASETS WITH PERSONAL DATA IN TANGENT	12
6 ETHICAL ASPECTS	13
7 CONCLUSIONS.....	14
8 REFERENCES.....	14
9 ANNEX 1 DATASET INFORMATION SHEET.....	15
10 ANNEX 2 LIST OF TANGENT DATASETS	16

List of tables

Table 1: Data manager.....	12
Table 2: Datasets with personal data.....	13

1 List of abbreviations and acronyms

Acronym	Meaning
EC	European Commission
GA	Grant Agreement
WP	Work Package
DMP	Data Management Plan
DM	Data Manager
IPR	Intellectual Property Rights
GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Interoperable, Reusable
UK	United Kingdom
EU	European Union
MB	Megabyte
PC	Project Coordinator

2 Introduction

This document describes the Data Management Plan (DMP) for the TANGENT project. The aim of this document is to provide an overview of the main elements of the data management policy that are going to be used by the consortium in this project, with regard to all the datasets that will be generated, harvested and/or used by the project. The DMP has the following characteristics:

- It covers the entire project life cycle, including how the information collected during the project will be preserved.
- it is not a static document; it evolves and gains more precision and substance during the lifespan of the project therefore other updated versions will be prepared at least at M18 and M36.

TANGENT project is part of the Open Research Data Pilot initiative managed by the EC, which aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.

TANGENT partners are committed to offer as much information as possible generated during the project through Open Access. However, TANGENT works with sensitive industry data from partners and 3rd parties, which cannot be made publicly available – but the listed principles can be applied for data sharing between the partners and inside the consortium, where applicable and necessary.

In the case of personnel data management, the consortium will comply with the requirements of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation - GDPR) and “UK - Data Protection Act 2018” on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The procedures that will be implemented for data collection, storage, and access, sharing policies, protection, retention and destruction will be according to the requirements of the national legislation of each partner and in line with the EU standards. For this reason, in case of personal data collection and processing, only anonymous user data will be collected and securely stored. Anonymous identification of user-provided information will be leveraged only to confirm the authenticity of users interacting with the system and to prevent malicious behaviour. No need to personally identify users through their information is envisioned nor to include sensitive data. The collected data will be treated anonymously and additionally a various set of measures will be put in place in order to protect user privacy and its data security, by embedding privacy by design principles from the early stage of the project technical start.

2.1 FAIR data

TANGENT follows the EC Guidelines for Open Access as well as the principles of FAIR data, this means that data is findable, accessible, interoperable, and reusable. In the next paragraph we provide a short description of each component¹:

- Data are **Findable** when they are described by sufficiently rich metadata and registered or indexed in a searchable resource that is known and accessible to potential users. Additionally,

¹ European Commission, Turning FAIR into reality

a unique and persistent identifier should be assigned such that the data can be unequivocally referenced and cited in research communications. The identifier enables persistent linkages to be established between the data, metadata and other related materials in order to assist data discovery and reuse. Related materials may include the code or models necessary to use the data, research literature that provides further insights into the creation and interpretation of the data and other related information.

- **Accessible** data objects can be obtained by humans and machines upon appropriate authorisation and through a well-defined and universally implementable protocol. In other words, anyone with a computer and an Internet connection should be able to access at least the metadata. It is important to emphasise that Accessible in FAIR does not mean Open without constraint. Accessibility means that the human or machine is provided - through metadata - with the precise conditions by which the data are accessible and that the mechanisms and technical protocols for data access are implemented such that the data and/or metadata can be accessed and used at scale, by machines, across the web.
- **Interoperable** data and metadata are described in the FAIR principles as those that use a formal, accessible, shared, and broadly applicable language for knowledge representation. They use vocabularies which themselves follow the FAIR principles, and they include qualified references to other data or metadata. What this describes is semantic interoperability. In other words, the data are described using normative and community recognized specifications, vocabularies and standards that determine the precise meaning of concepts and qualities that the data represent.
- For data to be **Reusable**, the FAIR principles reassert the need for rich metadata and documentation that meet relevant community standards and provide information about provenance. This covers reporting how data was created (e.g. survey protocols, experimental processes, information about sensor calibration and location) and information about data reduction or transformation processes to make data more usable, understandable or 'science-ready'. Reusability also requires that the data be released with a 'clear and accessible data usage license': in other words, the conditions under which the data can be used should be transparent to both humans and machines.

3 Dataset structure in TANGENT

3.1 Purpose of use of data in TANGENT

TANGENT aims to develop new complementary tools for optimising traffic operations from a multimodal perspective, both for passenger and freight transport. The project allows the precise knowledge and management of the mobility flows among the transport modes and enables the implementation and integration of innovative mobility solutions, services, and business models.

The TANGENT system will collect and process data from various sources to deliver these information services. Moreover, the TANGENT system will generate new datasets related to traffic optimisation, prediction, and users' behaviors. We report a preliminary list of datasets collected by TANGENT. The final list and the new datasets produced will be reported in the following versions of the Data Management Plan.

3.2 Preliminary list of dataset collected by TANGENT

Task 2.1 of the TANGENT project aims to collect available open data and private data sources with the direct involvement of data providers for TANGENT Case Studies (WP7). Particular attention will be 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 data collection activity is ongoing, and the final list of TANGENT datasets is planned for M12.

The current list of TANGENT datasets is reported in Annex I, where each dataset is described using the following metadata:

- **Identifier:** a unique identifier of the dataset (see Section 4.1);
- **Description:** short text describing the content of the dataset;
- **Type:** specifying if the dataset contains *static*, *real-time*, or *historical* data;
- **Format:** in which format (e.g., .csv, .json) the dataset can be obtained;
- **Size:** the size in Megabyte (MB) of the dataset;
- **Provider:** the data source of the dataset;
- **Accessibility:** specifying if the dataset is *open* or *private*;
- **Link:** a unique reference to access/download the dataset.

4 Data management in TANGENT

Data generation and management is a complex and prolonged process that requires multi-stakeholder bodies to work together in an interdisciplinary manner. For this, every partner will appoint a Data Manager (DM) who will be responsible for the management and control of the data sets retrieved and handled by the partner. The DM of the data provider will complete a template for each of the data sets generated. The template is available in Annex 1 of this document.

For each data set collected, processed and/or generated in the project, the following elements are considered:

- Dataset identifier
- Metadata profile
- Secure data sharing
- Archiving and preservation

4.1 Dataset Identifier

Unique identification of datasets is ensured by following provisioned unique naming convention drafted for the purpose of the TANGENT project. The convention for the dataset naming is as follows:

- Three capital letters to identify the demo site the dataset refers to;
 - “REN” is used for Rennes
 - “LIS” is used for Lisbon
 - “ATH” is used for Athens
 - “MAN” is used for the Greater Manchester
- Four capital letters to identify the subject (e.g., transportation mode) the dataset 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
- “PEDE” means pedestrian
- “OTHD” means other data
- “DEMR” means demand responsive transportation mode
- “TRAF” means traffic data
- Two digits to indicate a progressive number of the dataset.

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

4.2 Metadata Profile

Data collected, processed, or generated within the project will have its description to explain the dataset in more detail. A specific metadata profile will be defined and adopted for this purpose. In compliance with ongoing initiatives and projects (e.g., the NAPCORE project), DCAT-AP² version 2.0.0 has been selected as the starting point to define the metadata profile. An extension of DCAT-AP will be produced if needed to cover the specificity of TANGENT datasets. The definition of the metadata profile is an ongoing activity related to WP2 Task 2.2, and the profile will be included in D2.2 (M14). The data needed to fill the profile will be provided by the data owner/producer and/or other stakeholders and complemented with context (location, date, time) and publicly available information.

4.3 Secure data sharing

The TANGENT project will define how data will be shared and more specifically the access procedures, the embargo periods, the necessary software and other tools for enabling re-use, for all datasets that will be collected, generated, or processed in the project.

In case the dataset cannot be shared, the reasons for this will be mentioned (e.g. ethical, rules of personal data, intellectual property, and commercial, privacy-related, security-related).

In addition, beneficiaries do not have to ensure open access to specific parts of the research data if the achievement of the action's main objective, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management will present the reasons for not giving access.

4.4 Archiving and preservation

The data sharing procedures will be different across the datasets depending on the license and will be in accordance with the Grant Agreement.

The identification and analysis of data sources is in progress and the data anonymisation technique will be selected after this analysis.

Appropriate technical and organizational measures will be taken against unauthorised or unlawful processing of personal data in order to ensure that the individual cannot be identified from the captured data, and furthermore, we will ensure that data that could eventually lead to subsequent determination of the individual's path (where, when, how fast) will be stored for maximum of 24 hours. This way the possibility of identification of individuals is sufficiently minimized. The system will aggregate the data collected in short intervals which will ensure the anonymity of the data.

² <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/distribution/dcat-ap-200-pdf>

4.5 Data manager

During the project, each DM will be responsible for:

- completing the template in Annex 1 for the dataset providing and sending to the project coordinator
- sending the public datasets (in a raw and processed way if possible) to the project coordinator for uploading them in Zenodo
- sending all public documents, conference presentations and grey literature to POLIS for uploading them in Zenodo and to any other repository where it should be (for example, institutional repositories).

The Project Community page is already opened at Zenodo³ and will be curated by the Project Coordinator (PC). The PC will collect all templates generated during the course of the project, which will be stored until the final version is incorporated in the latest version of the DMP. The partners' DMs will be responsible for carrying out updates for each data template and for linking their updates to each template and also to the dataset.

Appointed Data Managers	
1. DEUSTO	Data manager: Leire Serrano (leire.serrano@deusto.es)
2. AIMSUN	Data manager: Athina Tympakianaki (athina.tympakianaki@aimsun.com)
3. NTUA	Data manager: Eleni Mantouka (elmant@central.ntua.gr)
4. IMEC	Data manager: Ynte Vanderhoydonc (ynte.vanderhoydonc@imec.be)
5. CEFRIEL	Data manager: Marco Comerio (marco.comerio@cefriel.com)
6. RUPPRECHT	Data manager: Morgane Juliat (m.juliat@rupprecht-consult.eu)
7. ID4CAR	Data manager: Véronique Rottier (veronique.rottier@id4car.org)
8. RENNES	Data manager: Pierre Renault (pi.renault@rennesmetropole.fr)
9. A-to-Be	Data manager: Lara Moura (lara.moura@a-to-be.com)

³ https://zenodo.org/communities/tangent_h2020/

Appointed Data Managers
10. CARRIS Data manager: João Vieira (joao.vieira@carris.pt)
11. TFGM Data manager: Hanna Tune (hannah.tune@tfgm.com)
12. PANTEIA Data manager: Maria Rodrigues (m.rodrigues@panteia.nl)
13. POLIS Data manager: Emiliya Kamenova(EKamenova@polisnetwork.eu)

Table 1: Data manager

5 Datasets with personal data in TANGENT

All the activities executed within TANGENT project will comply with the easy regulation at EU and national level on privacy and data protection transparency. The project will manage several sources of personal data. In general, personal data will be kept at the source and only pseudo anonymized data will be shared among the partners ensuring full compliance with GDPR. Access to the personal data will be minimized and controlled using the methods already in place. Consent forms, when needed, have been distributed and collected (Further detail in D10.1).

Finally, before publishing any data fulfilling the obligations of the partners with respect to the Article 29 of the Grant Agreement, a screening by all the involved parties will be carried out in order to not disclose any personal or confidential information. Special care will be put in ensuring that no de-anonymization could be performed on the published data. If such a case were detected (or suspected), several actions like the deletion of the relevant data or aggregation at an upper level will be discussed and agreed before publishing the dataset.

The following table contains the data sources already identified:

Datasets	Source	Personal data	Users	Protection measure
Interviews	Microsoft office format (Word, excel, power point) or Google documents, excel, presentations, forms	Yes	Access to personal data only Rupprecht.	Controlled access
Surveys	Coney tool	Yes	CEFRIEL, NTUA	Controlled access

Datasets	Source	Personal data	Users	Protection measure
Contact lists of stakeholders forum	Parties that showed interest on the project by signing the Letter of Support	Yes	Rupprecht, POLIS, ID4CAR, Deusto, CEFRIEL	Controlled access
Historical tracking information	Google timeline	Yes	NTUA	Controlled access

Table 2: Datasets with personal data

6 Ethical aspects

In TANGENT the sensible data will be harvested, stored and processed compliance to ethic guidelines. This is detailed in the deliverables from “WP10 – Ethics requirements”:

As well as potentially also personal data will be processed the project has established a separate workpackage (WP10 - Ethics requirements) to tackle ethical issues. The 'ethics requirements' that the project must comply with are included as deliverables in this work package, see as follows:

D10.1. H - Requirement No. 1 (M6): The confirmation that the templates of the informed consent/assent forms and information sheets (in language and terms intelligible to the participants) are kept on file is provided. The confirmation that the copies of opinions/ approvals by ethics committees and/or competent authorities for the research with humans have been obtained and are kept on file will be provided.

D10.2. POPD - Requirement No. 2 (M6): The beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the ‘data minimisation’ principle). The report will cover the following:

- A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants.
- A description of the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing must be submitted as a deliverable.
- Description of the anonymisation/pseudonymisation techniques.
- A confirmation that detailed information on the informed consent procedures in regard to data processing is kept on file.
- A confirmation that templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) are kept on file.
- As the research involves tracking, an explanation how the data subjects will be informed of the existence of tracking, its possible consequences and how their fundamental rights will be safeguarded.

D10.3. EPQ - Requirement No. 3 (M6): The applicant must demonstrate that appropriate health and safety procedures conforming to relevant local/national guidelines/legislation are followed for staff and participating citizens involved in this project.

Furthermore, in the scope of “WP9 – Project management” there will be two deliverables that will follow up the progress of ethics activities in the project:

- **D9.4. Ethics monitoring. First release (M12):** The document will contain all the ethical requirements and issues arise in this project and the processes carried out to solve them.
- **D9.10. Ethics monitoring. Second release (M36):** The document will contain all the ethical requirements and issues arise in this project and the processes carried out to solve them.

7 Conclusions

The TANGENT project makes use of data from different nature, meaning, closed – shared – open data with main attributes: volume, velocity and variety. Data comes from different sources like open data sources, consortium members and also third parties in the course of the case studies realisation. Parts of the data are sensible data and potentially even personal data and thereby secure data management / sharing is an important issue to be tackled by the project. This will be taken into account from a technical as well as from an organisational viewpoint.

This Data Management Plan on hand is created as a living document that is maintained over time following the dynamic requirements of the TANGENT project and it acts as a guideline for the whole consortium in regards of any data management in the project.

8 References

1. European Commission (2016). Open access data management - H2020 online manual. Retrieved from Data management - H2020 Online Manual
2. European Commission (2018), Turning FAIR into reality

9 Annex 1 Dataset information sheet

- **Name of the dataset:** Name of the dataset.
- **Dataset provider:** Person or institution responsible for its collection, curation and storage. If different persons or institutions are responsible for different steps of the data management process, please include here all of them along each responsibility.
- **Dataset owner:** Owner of the data (might not be the provider).
- **Description:** Brief description of the data features and the purpose of the data
- **Format (Media type):** Doc, pdf, api, json, xml
- **License:** The license of the dataset or a link to it.
- **Language:** ISO code of language
- **Ethics and legal aspects:** Any ethics and legal aspects of the data contained must be indicated.
- **Technical aspects:** Any technical difficulty or known problem of the dataset must be indicated.
- **Other:** Any other comment about the dataset must be included.

10 Annex 2 List of TANGENT datasets

The following table shows the current list of TANGENT datasets. The list is currently unbalanced towards the cities involved in the project (Rennes, Lisbon, Athens, and the Greater Manchester) as it reports the preliminary data collection performed by WP2. Data collection is a work in progress activity, and its finalization is scheduled for M12.

Identifier	Description	Type	Format	Size	Provider	Accessibility	Link
MAN_CARS_01	Count of cars across all the countlines from the sensors the user has access to.	Real time/ Historical	Json	<1Mb	VivacityLabs	Private. Authorized access through API	M_CARS_01
MAN_PEDE_01	Count of pedestrian across all the countlines from the sensors the user has access to.	Real time/ Historical	Json	<1Mb	VivacityLabs	Private. Authorized access through API	M_PEDE_01
ATH_CARS_01	Athens road network	Static	osm	1.3Mb	Open Street Map	Open data	A_CARS_01
REN_BUST_01	Station topological map	Static	Json/csv	0.79Mb	Rennes Metropole	Open data	R_BUST_01
REN_BUST_02	Bus routes of the STAR network	Static	Json/csv	5.5Mb	Rennes Metropole	Open data	R_BUST_02
REN_BUST_03	Real time buses position	Real time/ Historical	Json/csv	0.2Mb	STAR Data Explore	Open data	R_BUST_03
REN_BUST_04	Geolocation of the STAR network buses	Real time/ Historical	Json/csv	0.2Mb	STAR Data Explore	Open data	R_BUST_04
REN_BUST_05	Bus services	Static	Json/csv	2.75Mb	Rennes Metropole	Open data	R_BUST_05
REN_BUST_06	Physical stops of the STAR network	Static	Json/csv	1Mb	Rennes Metropole	Open data	R_BUST_06
REN_BUST_07	Logical stops of the STAR network	Static	Json/csv	0.6Mb	Rennes Metropole	Open data	R_BUST_07
REN_BUST_08	Real-time traffic alerts on STAR network lines	Real time/ Historical	Json/csv	0.4Mb	Rennes Metropole	Open data	R_BUST_08
REN_BUST_09	STAR line performance: max attendance level	Static	Json/csv	2.6Mb	Rennes Metropole	Open data	R_BUST_09

Identifier	Description	Type	Format	Size	Provider	Accessibility	Link
REN_BIKE_01	Localisation of STAR bike stations	Static	Json/csv	0.07Mb	Rennes Metropole	Open data	R_BIKE_01
REN_BIKE_02	Topology of bike-share stations	Static	Json/csv	0.02Mb	Rennes Metropole	Open data	R_BIKE_02
REN_BIKE_03	Real time station status	Real time/ Historical	Json/csv	0.03Mb	Rennes Metropole	Open data	R_BIKE_03
REN_BIKE_04	Cycle counts	Real time/ Historical	Json/csv	38.4Mb	Rennes Metropole	Open data	R_BIKE_04
REN_BIKE_05	Pedestrian counts	Real time/ Historical	Json/csv	38.4Mb	Rennes Metropole	Open data	R_BIKE_05
REN_BIKE_06	Cycling facilities and traffic calming zones	Static	Json/csv	5.6Mb	Rennes Metropole	Open data	R_BIKE_06
REN_METR_01	Schedules and lines for metro and underground	Static	Json/csv	0.3Mb	Rennes Metropole	Open data	R_METR_01
REN_METR_02	Status of the STAR metro lines network	Real time/ Historical	Json/csv	0.01Mb	Rennes Metropole	Open data	R_METR_02