ITN-5VC

Integrated Telematics for Next Generation 5G Vehicular Communications

ITN-5VC D5.3

Data management plan

Version v1.0 Date: 2020/05/23

Document properties:

Grant Number:	955629
Document Number:	D5.3
Document Title:	Data management plan
Partners involved:	UPV
Authors:	Danaisy Prado, Jose F. Monserrat
Contractual Date of Delivery:	2020/03/31
Dissemination level:	PU[1]
Version:	1.0
File Name:	ITN-5VC D5.3_v1.0

[1] CO = Confidential, only members of the consortium (including the Commission Services)

PU = Public

Table of Contents

Revision history

Version	Publication Date	Changes	Contributors
V0.1	2021/03/15	Initial version	Jose F. Monserrat (UPV)
V1.0	2021/05/23	Final version ready for Project Review	Danaisy Prado (UPV)

Executive summary

This document describes the Data Management Plan (DMP) for the ITN-5VC project, covering all four years (2020-10-01 until 2024-09-30). The DMP will serve as a guide for the management of the research data that will be handled in the frame of the project. This DMP details the procedures to follow in the acquisition/generation and storage of data. Additionally, the access and sharing policies will be discussed under the FAIR data principles. Security and ethical aspects will be also addressed in this document. During the life of the project, this document will be updated whenever the research activity requires it.

List of acronyms and abbreviations

DMP	Data management plan
EC	European Commission
FAIR	Findability, Accessibility, Interoperability and Reusability
GDPR	General Data Protection Regulation
UPV	Universitat Politècnica de València
WP	Work Package
5G	Fifth Generation

1 Introduction

1.1. Scope and objectives of the document

The Data Management Plan (DMP) describes the data management life cycle for the data to be handled within the ITN-5VC project. The DMP aims at settling a data management strategy to accomplish with the data principles: findability, accessibility, interoperability, and reusability (FAIR). At the same time, special attention should be paid to the data security and ethical issues, then, in this document those topics are also addressed.

1.2. Structure of the deliverable

The structure of this document follows the guideline provided by the H2020 program on FAIR Data Management.

- Section 1: Introduction
- Section 2: Data management in ITN-5VC
- Section 3: FAIR Data
- Section 4: Allocation resources
- Section 5: Data security
- Section 6: Ethical aspects
- Section 7. Conclusions

2 Data management in ITN-5VC

2.1 Data Summary

Being in line with the H2020 guidelines regarding the DMP, the main objective of this section is to give clear definitions of the data characteristics in order to ensure their understanding by the partners of the consortium.

Purpose of data collection/generation

The main purposes of data collection/generation identified in the ITN-5VC project so far are the following:

- Communication and Dissemination of project outcomes
- Contributions to deliverables
- Members' contact data to be included on publications
- Collection of geolocation information coming from the vehicles
- Project web page and repository log in

Types of data

The data that will be generated and processed in this project can be classified as follows:

- Project deliverables
- Articles published in scientific journals
- Articles presented in Conference and Workshops
- Measurement reports
- Presentations
- Meeting minutes

Formats of data

The formats in which the date will be stored are:

- Text: PDF, docx, xlcs, csv, pptx, xls
- Images: jpg
- Videos: MPEG2, MP4
- Audio: MP3, MP4

Origin of data

The data will come from partners.

Expected size of data

The work package leaders should contribute to define the expected size of data that will be handled in the project. At this point of the project the estimation of the size of the data cannot be given, however this information will be included in future updates of the DMP.

Data utility

The data generated during the execution of the project will be useful for the overall research community and foreseeably more interesting for those whose lines of research are focused on vehicular communications in the scope of 5G technologies.

3 FAIR data

According to the H2020 guidelines regarding to data management, the ITN-5VC project guarantees that the data follow the FAIR principles.

3.1 Findability

Metadata will be added to the research data in order to facilitate their findability and reusability. This way, other researchers can easily determine whether the dataset is relevant for their research. Metadata will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

A unified convention for naming and version handling of files is used in the project. For deliverables and internal technical reports, the following naming is used:

- Deliverables have names according to the following template “ITN-5VC-DX.Y-vN.M” where X.Y is the deliverable number and N.M indicates the version number as explained in the following section.
- Reports follow the same naming convention as deliverables, but with DX.Y replaced by IRX.Y.
- For journals, articles, and standard contributions, the naming is as follows: <Event>_<yyyy>_<Aut>_<Words> _<vN.M>, with the following explanation:
- <Event> - abbreviation of an event name such as a magazine, journal, conference, e.g., VTC, 3GPP, etc.
- <yyyy> - year of the publication, e.g., if the article was published in 2011 then indicate 2011
- <Aut>- first three letters of the last name of the author; in case of several authors to the paper, indicate only the last name of the first author and append the rest with “etal”, e.g. if the author is Tom Smith, the abbreviation would be Smi; in case of several authors, e.g., Tom Smith and Robert Brown, the abbreviation would be Smi_etal

- <Words> – use two meaningful words from the title of the publication, indicating what the publication is about, but not longer than ten letters in total; if it is longer, shorten it to 8 letters, e.g., if the full title of a publication is “System Performance of MIMO in WCDMA”, the meaningful words would be “MIMO_WCDMA”
- <vN.M> for version handling.

All documents follow the same rules for version handling:

- All versions before submitting to the EU commission start with N = 0, e.g., “v0.1”, “v0.2”, etc.

All versions submitted to the EU commission have M = 0. This means that, for instance, “v1.0” denotes the first submitted version, and “v2.0” the second submitted version.

3.2 Accessibility

In order to maximize the impact of the project and in accordance with the Horizon 2020 Open Research Data Pilot in which this consortium participates, the results will be shared with the research community through scientific publications in journals, conferences, workshops and other dissemination events. In addition, the results will be made available in open access data repositories. If access to certain data is restricted, adequate justification must be provided.

3.3 Interoperability

The exchange and re-use of the data and metadata among partners should be assisted by vocabularies and shared knowledge representation codes for semantic consistency among researchers. Information on data and metadata vocabularies, standards, and methodology to follow to facilitate interoperability should be updated as the project progresses.

3.4 Reusability

In order to guarantee the reusability principle, the data and metadata should be properly described. This implies that the origin of data and metadata should be precisely mentioned and data and metadata should comply with the relevant community standards for the sector. Additionally, clarifying licensing is needed for data to be as much reusable as possible.

4 Allocation of resources

For data to be FAIR, certain measures must be taken, the cost of which is unknown at the present phase of the project. However, these costs can be perceived as eligible costs. The budget for dissemination activities will be allocated among the ITN-5VC members who aim to issue scientific publications. The allocation of these resources should be sufficient to accommodate, at least partially, the requirements of the project for making data FAIR. The costs for data preparation and management will also be part of the project’s expenses.

5 Data security

Under the scope of ITN-5VC project, all members are committed to:

- Store data in at least two separate locations to avoid loss of data
- Encrypt data if it is deemed necessary
- Limit the use of USB flash drives
- Label files in a systematically structured way in order to ensure the coherence of the final dataset

All project deliverables and data will be stored and shared in the project repository restricted to the project consortium. This repository is hosted in the project web page server that uses HTTPS protocol and can be accessed through the web page using the corresponding credentials. Once, the deliverables have been reviewed by the all the partners, these could be put in open access, considering its dissemination level, with a disclaimer clarifying its review and acceptance states by the EU commission. Additionally, scientific publications and articles will be available in open source repositories to promote the data making FAIR.

6 Ethical aspects

ITN-5VC fully intends to address the ethical issues on data protection and decision-making processes related to this project, which it will do in the following manner:

Data Protection

Autonomous vehicles will require of geolocation for their appropriate functioning and integration in smart cities. These aspects imply that the providers of geolocation services may extract patterns and habits from users and deduce from this specific profiles or activities. The ITN-5VC project will comply with the current GDPR. All the collected data in the ITN-5VC project will be appropriately anonymized, so that no information of individuals may be extracted in order to identify them.

Decision making processes

There are ethics issues related to the capability of intelligent algorithms to make decisions, which may involve the way the autonomous vehicle reacts in an unpredicted situation such as trying to avoid an accident, swerving an animal on the road, avoiding a parked animal or, most controversially, having to decide whether to take a decision that could hurt either a passenger or someone outside it. In the meantime, until further research outputs are decided upon and legislation of the EC agreed on, ITN-5VC will consider the solution which many other automated cars manufacturers are implementing at the moment, which is for the fully automated car to stop when faced with an uncertain situation as described above.

In the ITN-5VC project, no additional ethical issues are addressed. In order to get more detailed information, see Section 4 of the ITN-5VC Project Grant Agreement.

7 Conclusions

In this document the bases and principles for the data management within the ITN-5VC project have been described. The general and particular aspects of each stage of the data management cycle have been addressed, serving as a basis for outlining the project's data management strategy. This is a living document, that will be updated during the life of the project according to the scientific activity developed.

Appendix A

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions

2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

HISTORY OF CHANGES		
Version	**Publication date**	**Change**
1.0	13.10.2016	▪ Initial version