Grant Agreement N°: 952189

Topic: ICT-53-2020

Next generation connectivity for enhanced, safe & efficient transport & logistics

D2.2: Data Management Plan

1st release

Revision: v.1.1

Work package	WP 2
Task	Task 2.3
Due date	28/2/2021
Submission date	24/12/2021 (resubmission after Interim Technical Review)
Deliverable lead	Dutch Ministry of Infrastructure and Water Management
Version	1.1

Abstract

This document is the first version of the 5G-Blueprint Data Management Plan (DMP). The document is set up to constitute a living document and intended to be complemented at a finer level of granularity through successive updates as the implementation of the project progresses and specific details become available. As such the document will gain more precision & substance over the lifespan of the project.

Keywords:

Data Management, DMP, GDPR

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	8/2/2021	First draft	Wim Vandenberghe (MIW)
V0.2	12/02/2021	Reviewed	Eric Kenis (MOW); Sander Maas (Sentors)
V0.3	23/2/2021	Reviews processed	Wim Vandenberghe (MIW)
V0.4	24/2/2021	Reviewed	Johann Marquez-Barja (imec)
V1.0	25/2/2021	Final version	Wim Vandenberghe (MIW)
V1.1	24/12/2021	Revised version based on comments received during the Interim Technical Review meeting with the European Commission.	Najmeh Masoudi (Seafar), Joost Vandenbossche (Be-Mobile), Free Bruneel (Be-Mobile), Oliver Held (Roboauto), Rakshith Kusumakar (V-Tron), Tom Van de Ven (Locatienet), Dries Naudts (imec), Vasilis Maglogiannis (imec)

Project co-funded by the European Commission under H2020-ICT-2018-20		
Nature of the deliverable:	R	
Dissemination Level		
PU	Public, fully open, e.g. web	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to 5G-Blueprint project and Commission Services	√

* R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
OTHER: Software, technical diagram, etc

EXECUTIVE SUMMARY

This document is the first version of the 5G-Blueprint Data Management Plan (DMP), defined at the end of M06. It is based on the Template Horizon 2020 Data Management Plan[1]. According to the principles outlined in that template, and taking into account that at this moment in time the detailed requirements and technical architectures still are being analysed and defined, this first version of the DMP is not yet providing detailed answers on all questions listed in the DMP template. Instead, it is intended to constitute a living document in which information can be made available on a finer level of granularity through successive updates as the implementation of the project progresses and when significant changes occur, gaining more precision and substance during the lifespan of the project.

To validate the DMP framework presented in this deliverable, some first reflections regarding the type of datasets have been worked out in a first draft version of the 5G-Blueprint Dataset Register. This register includes entries for all 4 use cases and all 8 enabling functions that are to be piloted in the project. It also contains a registry for the network measurements that will be part of the pilots (WP5), the surveys that will be realized in the context of the governance and business modelling (WP3), and the list of contact details that is collected as part of the newsletter registration process on the project website.

An important element significantly impacting the approach presented in this DMP is the fact that due to the strategic nature of the foreseen results part of the project, the partners decided to opt-out the Pilot on Open Research Data (ORD) in Horizon 2020. However, the project does commit to collaborating with external projects or initiatives as much as possible on the level of knowledge transfer. All this is reflected in the limited content of the present document that relates to the details of FAIR data (Findable, Accessible, Interoperable, Re-usable).

The last section of the document provides further details on the allocation of resources (including DPO contact details where needed), data security, and ethical aspects (be it that D1.1 “H – Requirement No. 1” equally is considered to constitute a valuable source of information on this matter).

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

BO	Business Objective
CA	Consortium Agreement
CAM	Cooperative Awareness Message
CPM	Collective Perception Message
CSV	Comma Separated Values
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
EF	Enabling Function
GA	Grant Agreement
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
ORD	Open Research Data
POPD	Protection of Personal Data
RO	Regulatory Objective
TO	Tele-Operated / Technical Objective
VAM	Vulnerable Road User Awareness message
UC	Use Case

1 INTRODUCTION

The present document is the first version of the 5G-Blueprint Data Management Plan (DMP), defined at the end of M06. It is based on the Template Horizon 2020 Data Management Plan. **According to the principles outlined in that template**, and taking into account that at this moment in time the detailed requirements and technical architectures still are being analysed and defined, this **first version of the DMP does not yet provided detailed answers to all the questions belonging to that DMP template.** Instead, it is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur, gaining more precision and substance during the lifespan of the project. Therefore, **this first iteration should be regarded as a document defining the DMP framework ruling** and providing a basis for all corresponding data management details, once relevant details are sorted out and data can be collected. However, as a start and for reasons of validation, some initial reflections on the type of datasets (to be collected, and to be processed) have been worked out in this deliverable. The corresponding **details will be completed and updated in the next iterations of this document**, but this way it is clear how the presented framework will need to be used, and that it is fit for purpose.

An important element that significantly impacts the approach presented in this DMP is the fact that due to the strategic nature of the expected results that will be developed within the project, the partners decided to opt-out the Pilot on Open Research Data (ORD) in Horizon 2020. More details are given in section 3

In line with the principles introduced here, and also with the plans outlined in the Grant Agreement of the project, subsequent releases of this living document will be delivered at M18 and M36 of the project, reflecting the output of the project as the project evolves.

2 DATA SUMMARY

In this section of the Data Management Plan, characteristics are provided on the individual datasets projected to be collected or generated within the project. It hence can be considered as the **5G-Blueprint Dataset Register**.

Table 1 defines the entries (in the register) that typically have to be answered, for any dataset type. This way, it can be guaranteed that for any dataset collected or generated by any of the work packages, all corresponding and required information systematically is captured, in a uniform manner.

Table 1: Template Dataset Register record

Dataset name	
Purpose of the data collection / generation	
Relation to the objectives of the project.	
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	
Pilot site where the data will be captured (if applicable).	
Types and formats of the data	
Will existing data be re-used and how?	
Origin of the data	
Expected size of the data (if known)	
To whom might it be useful ('data utility')?	
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage.[2]	
Description of potential personal data, such as faces and license plates in recorded videos[2]	
Is the data sensitive according to the GDPR[3]?[2]	
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data) [2]	
Usage of certified repositories for long term	

[2] This information would also be included in a GDPR data processing agreement regarding the personal data contained in the Dataset described in this Dataset Register record.

[3] According to https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en : The following personal data is considered 'sensitive' and is subject to specific processing conditions: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation.

preservation and curation	
Implemented anonymisation / pseudonymisation techniques.	
Privacy by design considerations[2]	
Data breach protocol (how to handle at incidents) [2]	
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	

The subsections of this section below elaborate in more detail on these entries, for each of the Use Cases (UC) and 'Enabling Functions' (EF) part of 5G-Blueprint. Since detailed requirements and the technical architecture of the solutions to be developed / piloted in the project still are being analysed, the corresponding content in this first iteration of the DMP may be limited to some first thoughts, in line with the currently known limitations.

Corresponding details may or most likely will have to be updated in a next iteration of the Data Management Plan, but this way the project's approach towards the Dataset Register at least can be made clear, and step by step validated as a suitable framework.

2.1 Pilot activities Use Cases

2.1.1 Common for all Use Cases

Some of the elements belonging to the different Dataset Register entries are identical for all records related to the Use Cases. In order to optimize the readability of this document, they are therefore presented here once.

Common for all UC	
Relation to the objectives of the project.	The dataset is required for realizing the following project objectives as defined in section 1.1 of part B of Annex 1 of the 5G-Blueprint Grant Agreement • TO2[4]: Tailor and implement the prototype of a tele-operated system. • TO4: Validation of the end-to-end tele-operated transport solution supported by 5G in real-life scenarios, including cross-border conditions.
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	All data is timestamped. This dataset is considered to be the minimum dataset that is needed to support its purpose. This data cannot be aggregated, it needs to be available on the individual vehicle or vessel and specific test execution level to allow the analysis of correct technical functionality.

[4] TO stands for Technical Objective

Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data will be destroyed 60 months after the project, once no further audits on the results of the project can be expected.

2.1.2 Pilot activities UC1

UC1	
Purpose of the data collection / generation	Validating that the developed use case “Automated Barge Control” functions well technically.
Relation to the objectives of the project.	See section 2.1.1
Relevance and accordance with the ‘data minimisation’ principle in the envisaged use.	Following data are captured, for both live piloting of the UC functionality, and later analyses of the corresponding technical KPI’s: • GPS positions (location, speed, heading) of the barge • Video streams of the camera’s installed on the barge, filming the environment surrounding the vessel. • Operator inputs in both supervisory and direct control mode Also see section 2.1.1
Pilot site where the data will be captured (if applicable).	Zelzate, Antwerp
Types and formats of the data	See section 2.1.1
Will existing data be re-used and how?	No
Origin of the data	Capturing devices are installed on a specific vessel that will be used in the daily operation of a shipping company, but data are collected solely during specifically planned test days. The crew of the vessel and the remote operator will be informed on specific test day(s), and logically need to agree the data are being recorded.
Expected size of the data (if known)	The TOV will send out its position with a frequency of 1 Hz. A single position enclosed in a CAM message will have a size of about 100 bytes, depending on the content of the CAM. This means for every hour of operation, around 360 MB of data will be stored. The video streams : their data rate is between 5Mpbs and 25 Mbps, corresponding with 2-11 GB video footage per hour of operation.
To whom might it be useful ('data utility')?	Participants of T4.2 (which also research this UC in T4.9 and T4.10): Seafar, Vtron, HAN-AR

	Analysis of BM & Governance (WP3)
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by …, on the following storage infrastructure: … (to be defined in D2.3). Access rights are arranged as follows: … (to be defined in D2.3). Also see section 2.1.1
Description of potential personal data, such as faces and license plates in recorded videos[2]	No video will be shared thus no personal data will be shared
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	All data transmission is done while keeping cyber security hazards in mind.
Usage of certified repositories for long term preservation and curation	Cloud services are used for repositories
Implemented anonymisation / pseudonymisation techniques.	Internal procedures are being developed for this
Privacy by design considerations	Internal procedures are being developed for this
Data breach protocol (how to handle at incidents)	Internal procedures are being developed for this
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	Internal procedures are being developed for this

2.1.3 Pilot activities UC2

UC2	
Purpose of the data collection / generation	Validating that the developed use case "Automated driver-in-loop docking functionality" functions well technically.
Relation to the objectives of the project.	See section 2.1.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	Following data are captured for both live piloting of the UC functionality, and later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the

	docking truck • Video streams of the camera's installed on the truck, filming the environment surrounding the vehicle. • Video streams of the camera's installed on the docking location, filming the truck and its surrounding environment. • Operator inputs when controlling the truck, and docking service inputs when the truck is docking automated. Also see section 2.1.1
Pilot site where the data will be captured (if applicable).	Vlissingen
Types and formats of the data	See section 2.1.1
Will existing data be re-used and how?	No
Origin of the data	Capturing devices are installed on a prototype truck and that will not be part of the daily operation of a transportation company, and on a mobile crane that will be used in daily operations of the beneficiary Verbrugge (VRBR). Data are collected solely during specifically planned test days. The truck driver and the remote operator will be informed on specific test day(s), and logically need to agree the data are being recorded.
Expected size of the data (if known)	The GPS positions will hold about 360 kilobytes per hour of operation. Operator/docking service inputs: < 900 MB per hour of operation, The video streams : 10-100GB per hour of operation
To whom might it be useful ('data utility')?	Participants of T4.3/T4.4/T4.5/T4.6 (which also research this UC in T4.9 and T4.10): Vtron, HAN-AR, VRBR Analysis of BM & Governance (WP3)
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	Data generated during the course of this project will be stored and backed up on the networks/research drives secured by the institution (HAN University of Applied Sciences). Additionally, a copy of the data will also be stored and backed up in Microsoft teams to facilitate collaboration with project partners. Microsoft teams is also fully supported by the IT department of the institution. Access is given only to UC related project partners.
Description of potential personal data, such as faces and license plates in recorded videos[2]	To be determined which measures are needed to avoid personal data in the video screens, e.g. by automated blurring of faces and license plates of passers-by.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on one specific pilot location, and the presence on that location is determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that

	could be derived from specific appearance characteristics of the participating individuals.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Default security measures of the institution networked research storage.
Usage of certified repositories for long term preservation and curation	Yes, we will use Microsoft Teams, and OneDrive.
Implemented anonymisation / pseudonymisation techniques.	N/A
Privacy by design considerations	N/A
Data breach protocol (how to handle at incidents)	Default security measures of the institution networked research storage.
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	No sensitive data will be collected as a part of this project. Security measures of the institution applies to all data generated in this project.

2.1.4 Pilot activities UC3

UC3	
Purpose of the data collection / generation	Validating that the developed use case "CACC based platooning" functions well technically.
Relation to the objectives of the project.	See section 2.1.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	Following data are captured, for both live piloting of the UC functionality, and later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the 2 cars forming a platoon • Video streams of the camera's installed on the cars, filming the environment surrounding the vehicle. • Operator inputs when controlling the truck, and docking service inputs when the truck is docking automated. • CACC function related data exchanged between the two vehicles in the platoon Also see section 2.1.1
Pilot site where the data will be captured (if applicable).	Zelzate, Antwerp
Types and formats of the data	The GPS positions, operator inputs and CACC data are collected in CSV/JSON/XML/other… (to be defined in D2.3) The video streams are collected in H.264/other (to be defined)?

Will existing data be re-used and how?	No
Origin of the data	Capture devices are installed on 2 prototypes that are not involved in daily operation of a transportation company. Data are collected solely during specifically planned test days. The car drivers and the remote operator will be informed on specific test day(s), and logically need to agree the data are being recorded.
Expected size of the data (if known)	Not stored, will be processed and discarded. In future, we will using the acceleration, GPS, speed, distance from lead, system status: active/not active, brake status, overrides, relative speed and distance from lead vehicle. Approximately. 0.4 KB/s
To whom might it be useful ('data utility')?	Participants of T4.7 (which also research this UC in T4.9 and T4.10): Vtron, HAN-AR, imec, TME Analysis of BM & Governance (WP3)
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	Locally as a FDR setup(new file every ride for approximately 15 days and then deleted) or in the local cloud in Vtron platform if necessary. Also see section 2.1.1
Description of potential personal data, such as faces and license plates in recorded videos[2]	To be determined which measures are needed to avoid personal data in the video screens, e.g. by automated blurring of faces and license plates of passers-by.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals.
Ideally we wouldn't store it in the cloud but only local in the vehicle.	Ideally we wouldn't store it in the cloud but only local in the vehicle.
Usage of certified repositories for long term preservation and curation	NA
Implemented anonymisation / pseudonymisation techniques.	Local data
Privacy by design considerations	NA
Data breach protocol (how to handle at incidents)	NA
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	NA

2.1.5 Pilot activities UC4

UC4	
Purpose of the data collection / generation	Validating that the developed use case “Remote take-over operations” functions well technically.
Relation to the objectives of the project.	See section 2.1.1
Relevance and accordance with the ‘data minimisation’ principle in the envisaged use.	Application log data. Also see section 2.1.1
Pilot site where the data will be captured (if applicable).	Vlissingen, Zelzate, Antwerp
Types and formats of the data	See section 2.1.1
Will existing data be re-used and how?	No
Origin of the data	Roboauto software.
Expected size of the data (if known)	The GPS positions and operator inputs: ~1kBps The video streams : 1-2MBps (depending on resolution, bitrate, number of video streams, etc.) This corresponds with roughly 3.5 – 7 GB per hour of operation.
To whom might it be useful ('data utility')?	Roboauto
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by Roboauto, on the following storage infrastructure: Roboauto cloud Access rights are arranged as follows: authorized Roboauto personnel.
Description of potential personal data, such as faces and license plates in recorded videos[2]	No videos being recorded.
Is the data sensitive according to the GDPR?	No.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Data is backed up every night to another Roboauto server at a different site using a secure connection. Access is controlled based on a certificate and password.
Usage of certified repositories for long term preservation and curation	None
Implemented anonymisation / pseudonymisation techniques.	None
Privacy by design considerations	No personal/sensitive data
Data breach protocol (how to handle at incidents)	None
Sensitive non-personal data but with potential commercial impact, such as business cases, expert’s personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	None

2.2 Pilot activities Enabling Functions

2.2.1 Common for all Enabling Functions

Some of the elements belonging to the different Dataset Register entries are identical for all records related to the Enabling Functions. In order to optimize the readability of this document, they are therefore presented here once.

Common for all EF	
Relation to the objectives of the project.	The dataset is required to realize the following project objectives as defined in section 1.1 of part B of Annex 1 of the 5G-Blueprint Grant Agreement • TO3: Implement and deploy enabling functions guaranteeing the safety of tele-operated transport • TO4: Validation of the end-to-end tele-operated transport solution supported by 5G in real-life scenarios, including cross-border conditions.
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	All data is timestamped. This dataset is considered to be the minimum dataset that is needed to support its purpose. This data cannot be aggregated, it needs to be available on the individual vehicle or vessel and specific test execution level to allow the analysis of correct technical functionality.
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data will be destroyed 60 months after the project, once no further audits on the results of the project can be expected.

2.2.2 Pilot activities EF1

EF1	
Purpose of the data collection / generation	Validating that the developed enabling function "Enhanced Awareness HMI" functions well technically.
Relation to the objectives of the project.	See section 2.2.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the teleoperated road vehicle • Information presented on the enhanced awareness dashboard (including corresponding location information): driving speed/speed advice, warning, navigation and routing features. • SRTI (Safety Related Traffic Information) data received as input data

	• Emergency vehicles location data received as input data (and originating from the existing Talking Traffic emergency vehicles warning service) • Data received from other EFs as input data
Pilot site where the data will be captured (if applicable).	Vlissingen, Zelzate, Antwerp
Types and formats of the data	The GPS positions are collected in json format (with WGS84 positions). They will be enclosed in CAM messages between EF's and in internal Be-Mobile format inside the EF. The information presented on the enhanced dashboard is presented in json or xml format, according to the Datex II protocol. The SRTI and emergency vehicles location data will be collected in the format as they are exposed by their corresponding source: DATEXII for SRTI and DENM messages for emergency vehicles.
Will existing data be re-used and how?	Yes: • the SRTI data will be sourced from the Data for Road Safety initiative (https://www.dataforroadsafety.eu/) • the emergency vehicles location data will be sourced from the corresponding Talking Traffic service that the "brancheverening Ambulancezorg Nederland" is providing (and which is implemented by Be-Mobile for this organisation)
Origin of the data	The existing data has the following origin: • SRTI data: connected vehicles sold by the vehicle OEMs participating in the data for road safety initiative, and for which the owners gave consent for sharing their SRTI data. 5G-Blueprint will only request so called L3 data from the data for road safety ecosystem, meaning that this data only refers to a specific observed traffic situation (e.g. slipper road detected at location X), and cannot be related to a specific vehicle (hence aggregated data). • Emergency vehicles location, driving direction and vehicle speed data: ambulances sharing their location information when driving with priority (and hence with active siren and flashing light), so that service providers can warn the neighbouring traffic about their arrival and the need to give way. The project-specific data is captured by devices installed on a prototype truck and two prototype cars that are not part of the daily operation of a transportation company. Data is only collected during specifically planned test days. The drivers and the remote operator are always aware if they are participating to a specific test day, and hence this data is being recorded.
Expected size of the data (if known)	The GPS positions will hold about 360 kilobytes per hour of

	operation. The information presented on the enhanced dashboard will depend on the amount of events and notifications that will be shown. The used SRTI feed has an average size of 900 kilobytes and 7.7 megabytes that is renewed every minute for Belgium and the Netherlands respectively, however not all of this data will be used, as most of it will be geographically irrelevant.
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: Be-Mobile
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data will be stored by Be-Mobile, on storage infrastructure inside the EC. Access rights are arranged as follows: every team member working on the project, including the developers as well as the testers and project manager.
Description of potential personal data, such as faces and license plates in recorded videos[2]	Does not apply in this EF.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	- data classification - access control (oauth 2.0, certificates, ip whitelisting, depending on data) - backups - mandatory secure protocols and use of encrypted channels
Usage of certified repositories for long term preservation and curation	Not applicable; the data will be deleted after completion of the project
Implemented anonymisation / pseudonymisation techniques.	- each session has a unique ID, exception for use cases where a fixed ID must be used. - when sending to 3rd parties, these sessionIDs are extra anonymized. - initial positions are not stored, except for some special use cases. This makes home-work determination even harder
Privacy by design considerations	- dpo determines DPA use for every project - DPIA is made if relevant - Security by design principles
Data breach protocol (how to handle at incidents)	There is an internal protocol available Be-Mobile, in which security officers of Be-Mobile are informed and they will

	give instructions on how to proceed.
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	Does not apply

2.2.3 Pilot activities EF2

EF2	
Purpose of the data collection / generation	Validating that the developed enabling function "Vulnerable Road User (VRU) interaction" functions well technically.
Relation to the objectives of the project.	See section 2.2.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the teleoperated road vehicles • GPS positions (location, speed, heading), VRU type (pedestrian or cyclist), sensory handset data (accelerometer, etc.), and planned route of Vulnerable Road Users.
Pilot site where the data will be captured (if applicable).	Vlissingen, Zelzate, Antwerp
Types and formats of the data	The GPS positions and other VRU characteristics are collected in CAM & VAM messages, collected in a database. The information exposed to the enhanced dashboard is collected in textfiles containing JSON.
Will existing data be re-used and how?	No
Origin of the data	To capture the VRU data, a specific application will be installed on the smartphone of specific test participants, which will always be employees of parties belonging to the project's group of beneficiaries, or to the project's advisory board, and which will never be children, adults unable to give consent, vulnerable individuals/groups, nor persons with disabilities or reduced mobility/orientation. Hence the VRU role in the tests of the pilots will always be well-instructed research participants which exactly know how they should behave within their role of cyclist or pedestrian in the test. Data is only collected during specifically planned test days. The drivers, remote operator and VRUs are always aware if they are participating to a specific test day, and hence this data is being recorded. No VRU (hence cyclists or pedestrians) will be tracked during their day-to-day movements.
Expected size of the data (if known)	The GPS positions and other VRU characteristics: 500 to 5000MB

	The information exposed to the enhanced dashboard: 500 to 5000MB
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: Locatienet
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by dedicated servers of Locatienet hosted by Info.nl in Amsterdam, The Netherlands. Access rights are arranged as follows: only personell of Locatienet involved in 5G Blueprint have access to the data. The project manager of LN administers the access rights.
Description of potential personal data, such as faces and license plates in recorded videos[2]	Location data
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Industry grade access control to the server and data store is implemented. Data access is restricted to personnel that is involved in the project.
Usage of certified repositories for long term preservation and curation	Not applicable; the data will be deleted after completion of the project.
Implemented anonymisation / pseudonymisation techniques.	Not applicable; all personal data will be collected by handsets operated by LN personnel. For the purpose of the evaluation each VRU has to be identifiable.
Privacy by design considerations	For the pilot the VRU need to be identifiable but the service can be used anonymously once operational. Geographic filtering in the MQTT service ensures only vehicles in the immediate vicinity of a VRU can retrieve location data on the VRU.
Data breach protocol (how to handle at incidents)	In the case of incidents the server is simply shut down and LN management is alerted.
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	NA

2.2.4 Pilot activities EF3

EF3	
Purpose of the data collection / generation	Validating that the developed enabling function "Time slot reservation at intersections" functions well technically.
Relation to the objectives of the project.	See section 2.2.1

Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the teleoperated road vehicles • Video streams of the camera's installed on the road vehicles, filming the environment surrounding the vehicle. • Information received from the intelligent traffic light (MAP with intersection topology, iSPAT with information on the reserved time slot), and information sent to the intelligent traffic light (CAM messages containing vehicle location, and SRM message requesting a time slot). • Information exposed to the enhanced awareness dashboard (and the corresponding location information).
Pilot site where the data will be captured (if applicable).	Vlissingen, Zelzate
Types and formats of the data	The GPS positions and intelligent traffic light data are collected in JSON To be further discussed: can video streams made available via the existing road surveillance camera's owned by North Sea Port? The information exposed to the enhanced dashboard is collected in JSON
Will existing data be re-used and how?	No
Origin of the data	The capture devices are installed on a prototype truck and two prototype cars that are not part of the daily operation of a transportation company. The traffic lights involved in the pilot will be upgraded to intelligent traffic lights, which will operate as production road infrastructure, in which the experimental time slot reservation function can be activated during specifically planned test days. Data is only collected during these test days. The drivers and remote operator are always aware if they are participating to a specific test day, and hence this data is being recorded.
Expected size of the data (if known)	< 1 Mbps (and hence < 450 MB per hour of operation). The GPS positions of the ToV are handled in the RSU for mapping aspects. The mapping is on the accuracy of lane level determined).
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: Swarco, Be-Mobile
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored in the RSU is based on a standard logging principle which is agreed in the Concorda project. The logging data is stored in a repository system. Access rights are arranged as follows: via the login procedure on the MyCity platform every person who is qualified by our authorization office, for which data access is checked and on which level access is permitted Also see section 2.2.1
Description of potential personal data, such as faces and license plates in	To be determined which measures are needed to avoid personal data in the video screens, e.g. by automated

recorded videos[2]	blurring of faces and license plates of passers-by, of course only in case of using roadside cameras of NSP (which is to be further discussed).
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	The access to the data is strictly regulated by our authorization office. This is according our ISO 27001 defined procedure
Usage of certified repositories for long term preservation and curation	(to be defined in D2.3)
Implemented anonymisation / pseudonymisation techniques.	(to be defined in D2.3)
Privacy by design considerations	(to be defined in D2.3)
Data breach protocol (how to handle at incidents)	(to be defined in D2.3)
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	(to be defined in D2.3) Possibly to be added to section 2.3 with a reference to this EF.

2.2.5 Pilot activities EF4

EF4	
Purpose of the data collection / generation	Validating that the developed enabling function "Distributed perception" functions well technically.
Relation to the objectives of the project.	See section 2.2.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the teleoperated road vehicles • Video streams of the camera's installed on the road vehicles, filming the environment surrounding the vehicle. • Video streams of the camera's installed as road infrastructure. • Information exposed to the enhanced awareness dashboard (and the corresponding location information).

Pilot site where the data will be captured (if applicable).	Vlissingen, Zelzate, Antwerp
Types and formats of the data	The GPS positions are collected in json format (with WGS84 positions). The video and point cloud streams are collected in H.264/AVC. The information exposed to the enhanced dashboard is collected in JSON.
Will existing data be re-used and how?	No
Origin of the data	The capture devices are installed on a prototype truck and two prototype cars that are not part of the daily operation of a transportation company. The roadside infrastructure cameras in the pilot will be deployed specifically for the project (no production road surveillance cameras will be used), and will only be activated during specifically planned test days. Data is only collected during these test days. The drivers and remote operator are always aware if they are participating to a specific test day, and hence this data is being recorded.
Expected size of the data (if known)	LiDAR point-clouds: between 20 Mbps and 100 Mbps GPS/GNSS positions data: 20kb every 0.1s Video stream: 1.5Mb max every 0.2s (this value is per image per stream, for distributed data fusion we will be having multiple streams on the ego-agent (TOV) it-self, and on neighbouring vehicles on the road that can broadcast sensor data) Point cloud stream: 2 Mb max every 0.2s (this value is per point cloud sweep per stream, for distributed data fusion we will be having multiple streams on the ego-agent (TOV) it-self, and on neighbouring vehicles on the road that can broadcast sensor data) Information exposed to the enhanced dashboard:
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: imec
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored in a centralized location, on the following storage infrastructure: … (to be defined in D2.3). The data will be destroyed 60 months after the project, once no further audits on the results of the project can be expected.
Description of potential personal data, such as faces and license plates in recorded videos[2]	To be determined which measures are needed to avoid personal data in the video screens, e.g. by automated blurring of faces and license plates of passers-by.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals. The only exception would be

	other road users being filmed by the roadside cameras. To avoid such uninformed road users to be capture on camera, the camera's will only be activated during test execution, and during test executions bystanders will be informed about the test in progress, and requested to continue their journey outside of the filmed area.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Data will not be stored for long term use. However, data is backed up to another storage disk, that will automatically be wiped after an agreed period of time (legal term of 30 days). Access to the disk is controlled based on a certificate and password.
Usage of certified repositories for long term preservation and curation	Not applicable; the data will be deleted after completion of the project
Implemented anonymisation / pseudonymisation techniques.	Each agent will broadcast its decoded location and intermediate representation that can only be decoded by our internally developed algorithm, which will guarantee anonymisation.
Privacy by design considerations	Not applicable (No personal/sensitive data)
Data breach protocol (how to handle at incidents)	Does not apply, as EF4 does not have direct link with the teleoperator, and the tele-operated agent is able to handle perception on its own if there is breach in the tele-network.
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	Does not apply

2.2.6 Pilot activities EF5

EF5	
Purpose of the data collection / generation	Validating that the developed enabling function "Active collision avoidance" functions well technically.
Relation to the objectives of the project.	See section 2.2.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the teleoperated truck • Video streams of the camera's installed on the truck, filming the environment surrounding the vehicle. • Sensor data from additional, dedicated sensors mounted on the truck (e.g. Lidar). • Information exposed to the enhanced awareness dashboard (and the corresponding location information).
Pilot site where the data will be captured	Vlissingen, Zelzate, Antwerp

(if applicable).	
Types and formats of the data	The GPS positions are collected in json format (with WGS84 positions). The video streams are collected in VP-8 The information exposed to the enhanced dashboard is collected in text format.
Will existing data be re-used and how?	No
Origin of the data	The capture devices are installed on a prototype truck and that is not part of the daily operation of a transportation company. Data is only collected during specifically planned test days. The drivers and remote operator are always aware if they are participating to a specific test day, and hence this data is being recorded. The data capturing will be executed by the Roboauto software.
Expected size of the data (if known)	The GPS positions and other sensor data: 2MBps The video streams: 1-2MBps (captured as a part of UC4) . This corresponds with roughly 3.5 – 7 GB per hour of operation. The information exposed to the enhanced dashboard: …(to be defined in D2.3)
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: Roboauto
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by Roboauto, on the following storage infrastructure: Roboauto cloud. Access rights are arranged as follows: authorized Roboauto personnel.
Description of potential personal data, such as faces and license plates in recorded videos[2]	No videos are being recorded.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on fixed trajectories, and the presence on those trajectories are determined by the professional activities of the corresponding individuals, not by their personal activities. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Data is backed up every night to another Roboauto server at a different site using a secure connection. Access is controlled based on a certificate and password.
Usage of certified repositories for long term preservation and curation	/
Implemented anonymisation / pseudonymisation techniques.	/
Privacy by design considerations	No personal/sensitive data
Data breach protocol (how to handle at	/

incidents)	
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	/

2.2.7 Pilot activities EF6

EF6	
Purpose of the data collection / generation	Validating that the developed enabling function "Container ID recognition" functions well technically.
Relation to the objectives of the project.	See section 2.2.1
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • Video streams of the camera's installed on the spreader of the mobile crane, looking downwards towards the containers that are loaded • Information exposed to the enhanced awareness dashboard (and the corresponding location information).
Pilot site where the data will be captured (if applicable).	Vlissingen,
Types and formats of the data	The video streams are collected in H.264 stream The information exposed to the enhanced dashboard is pushed as REST-based API in JSON format.
Will existing data be re-used and how?	No
Origin of the data	The capture devices are installed on a crane that is part of an operational crane. The cameras will be deployed specifically for the project (no production surveillance cameras belonging to the daily operation of the port will be used), and will only be activated during specifically planned test days. Data is only collected continuously.
Expected size of the data (if known)	The video streams is expected to be 3.6GB per day (assumption: 300 handlings x 4 cameras x 3MB/clip) The information exposed to the enhanced dashboard is ~10MB/day, as this only consists of REST-messages. The message contains the URL's to these videos so that they can be deeplinked in the browser - but only if that user is authorized to the underlying database..
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: Sentors
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by Sentors on the following storage infrastructure: Google Cloud within European datacenters. Access rights are arranged by email address and password. Browser sessions expire and regular log-in is required.

	Access rights are only provided to a few persons from the operational planning, and the data is deleted after 3 to 12 months (depending on what is agreed with the crane operator). The videos are stored for 3 to 12 months, to be decided by the crane operator. The only reason for storing the data, is to have the videos available when damage disputes occur, i.e. when a container is damaged. The older the video, the less useful the contents. Therefore the videos can safely be deleted after this period.
Description of potential personal data, such as faces and license plates in recorded videos[2]	The vast majority of videos will only be from containers, and only when the spreader is at most a few meters away from the roof of the container. This is a dangerous situation for persons to be standing next to. So it is unlikely that personal data will be included in these videos. Incidentally, it might happen that someone in the barge or at the truck might be captured in the video clip. However, that probability is deemed low, and if so, it will be from a top view.
Is the data sensitive according to the GDPR?	No, as explained in the previous row, the videos do not contain personal data. No video images will be recorded of the involved research participants, to avoid revealing racial or ethnic origin, or political opinions, religious or philosophical beliefs that could be derived from specific appearance characteristics of the participating individuals.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	This is described in more details in Sentors' existing standard Data Processing Agreement.
Usage of certified repositories for long term preservation and curation	The data is deleted within 3 to 12 months, depending on what is agreed with the crane operator. There is no other long-term storage involved. Incidental images are stored in the training set of Sentors, to further optimize the machine learning algorithms. However, these images are manually selected and all personal and sensitive data is removed from the images.
Implemented anonymisation / pseudonymisation techniques.	This is not used. The data is searchable by container number. The video collection is enormous (thousands of video clips per day).
Privacy by design considerations	This is described in the data processing agreement that Sentors agrees with the crane operator. In principe, most filtering is already done in the image recognition software, so that the system by design only stores images of containers. Also, if no containers are handled, there is no twistlock/unlock action. In those cases, no videos are stored anyway.
Data breach protocol (how to handle at incidents)	This is described in the data processing agreement that Sentors agrees with the crane operator.In principe, any breach will be discussed by Sentors with the crane operator, and depending on the nature of the breach, will Authorised Persoonsgegevens.
Sensitive non-personal data but with	The most sensitive part is when particular container is

potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	actively tracked and traced by a criminal organization, in the pursuit of drug transport or stealing of high-value goods. Based on our discussions with police forces and customs, in real-life this probability is much higher at seaports from certain countries. Further non-personal sensitive data is related to the type and number of containers that the crane handles. However, this information is to a large extend already known and publicly marketed at the container terminal website, in terms of the amount of container volume (TEU) they handle and the nature of their services (e.g. reefer spots and dangerous goods permits)

2.2.8 Pilot activities EF7

EF7	
Purpose of the data collection / generation	Validating that the developed enabling function "ETA sharing" functions well technically.
Relation to the objectives of the project.	See section 2.2.1, extended with: • BO2[5]: Commercial possibilities
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI's: • GPS positions (location, speed, heading) of the tracked trucks • ETA Information exposed to EF8 (and the corresponding location information).
Pilot site where the data will be captured (if applicable).	NA
Types and formats of the data	The GPS positions are collected in json format (with WGS84 positions). They will be enclosed in CAM messages between EF's and in internal Be-Mobile format inside the EF. The ETA information is collected in JSON.
Will existing data be re-used and how?	No
Origin of the data	The capture devices are installed on one or more trucks that are part of the daily operation of a transportation company (beneficiaries Verbrugge and Joosen). The drivers and remote operator are made aware that during their daily activities their data is being recorded (similarly to the track and tracing already in place to support the planning operations of the organisation).

[5] BO stands for Business Objective

Expected size of the data (if known)	The GPS positions will hold about 360 kilobytes per hour of operation. The ETA information exposed to other EF's will be a small json file of unknow size.
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: Be-Mobile, Verbrugge, Joosen
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data will be stored by Be-Mobile, on storage infrastructure inside the EC. Access rights are arranged as follows: every team member working on the project, including the developers as well as the testers and project manager
Description of potential personal data, such as faces and license plates in recorded videos[2]	To be determined which measures are needed to avoid personal data in the video screens, e.g. by automated blurring of faces and license plates of passers-by.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on trajectories determined by the professional activities of the corresponding individuals, not by their personal activities.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	- data classification - access control (oauth 2.0, certificates, ip whitelisting, depending on data) - backups mandatory secure protocols and use of encrypted channels (to be defined in D2.3)
Usage of certified repositories for long term preservation and curation	Not applicable; the data will be deleted after completion of the project.
Implemented anonymisation / pseudonymisation techniques.	- each session has a unique ID, exception for use cases where a fixed ID must be used. - when sending to 3rd parties, these sessionIDs are extra anonymized. - initial positions are not stored, except for some special use cases. This makes home-work determination even harder
Privacy by design considerations	- dpo determines DPA use for every project - DPIA is made if relevant - Security by design principles
Data breach protocol (how to handle at incidents)	There is an internal protocol available Be-Mobile, in which security officers of Be-Mobile are informed and they will give instructions on how to proceed.
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	Does not apply

2.2.9 Pilot activities EF8

EF8	
Purpose of the data collection / generation	Validating that the developed enabling function “Logistics chain optimization” functions well technically.
Relation to the objectives of the project.	See section 2.2.1, extended with: • BO2: Commercial possibilities
Relevance and accordance with the ‘data minimisation’ principle in the envisaged use.	The following data is captured, both for live piloting of the EF functionality, but also for later analyses of the corresponding technical KPI’s: • GPS positions (location, speed, heading) of the tracked trucks • Video and audio streams of the camera’s installed on the truck, filming the environment surrounding the vehicle when stationary on a buffer parking or other parking lot. • Video and audio streams of camera’s installed at the parking area of Verbrugge and Joosen. • Information created by the EF regarding buffer parking status, integrity of parked vehicle, etc. • Potentially: ETA and additional (forward looking) planning and routing details, copied from dispatch centres and/or additional dedicated sources
Pilot site where the data will be captured (if applicable).	Vlissingen, Antwerp
Types and formats of the data	Details about the collection of the GPS positions, video streams, audio streams and EF8 output will be defined in D2.3
Will existing data be re-used and how?	No
Origin of the data	The capture devices are installed on one or more trucks that are part of the daily operation of a transportation company (beneficiaries Verbrugge and Joosen). The drivers and remote operator are made aware that during their daily activities their data is being recorded (similarly to the track and tracing already in place to support the planning operations of the organisation). It will be made sure that when parked (especially overnight), audio and video is only recorded when the driver is not present in or around the vehicle.
Expected size of the data (if known)	The GPS positions: will be defined in D2.3 The video streams: will be defined in D2.3 The audio positions: will be defined in D2.3 The information created EF8: will be defined in D2.3
To whom might it be useful ('data utility')?	Participants of WP6 working on this EF: R40, Verbrugge, Joosen
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of	Details about data storage and access rights will be defined in D2.3.

storage	
Description of potential personal data, such as faces and license plates in recorded videos[2]	To be determined which measures are needed to avoid personal data in the video screens, e.g. by automated blurring of faces and license plates of passers-by.
Is the data sensitive according to the GDPR?	No, there is personal data, but it is not classified as sensitive data. Location data is only collected on trajectories determined by the professional activities of the corresponding individuals, not by their personal activities. To be determined which measures need to be defined to make sure that the video and audio captured when parked are also not sensitive.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Will be defined in D2.3 Will be defined in D2.3 Will be defined in D2.3
Usage of certified repositories for long term preservation and curation	Will be defined in D2.3
Implemented anonymisation / pseudonymisation techniques.	Will be defined in D2.3
Privacy by design considerations	Will be defined in D2.3
Data breach protocol (how to handle at incidents)	Will be defined in D2.3
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	Will be defined in D2.3

2.3 Pilot measurements network/connectivity

Network/connectivity (WP5)	
Purpose of the data collection / generation	Validating that the designed 5G network architecture to support all use cases and enabling functions is able to meet the imposed requirements.
Relation to the objectives of the project.	This dataset is required to realize the following project objectives as defined in section 1.1 of part B of Annex 1 of the 5G-Blueprint Grant Agreement • TO1: Design and implement a 5G network for CAM services • TO4: Validation of the end-to-end tele-operated transport solution supported by 5G in real-life scenarios, including cross-border conditions.
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	The following data is captured, both for live piloting of the UC and EF functionality, but also for later analyses of the corresponding technical KPI's: • Connectivity characteristics at the mobile device (end-to-end latency, service coverage, service continuity, service reliability, bandwidth capacity,

	resource allocation efficiency, …) • Network operation data at the network side (to be determined what this exactly will be) All this data is timestamped. This dataset is considered to be the minimum dataset that is needed to support its purpose. This data cannot be aggregated, it needs to be available on the individual vehicle and specific test execution level to allow the analysis of correct technical functionality.
Pilot site where the data will be captured (if applicable).	Vlissingen, Zelzate, Antwerp
Types and formats of the data	The data is collected in CSV format by each of the involved partner according to well-defined log formats (to be determined in WP5 T5.4). The logged data will then be imported into a central database.
Will existing data be re-used and how?	No
Origin of the data	On the three pilot sites, the Mobile Network Operators Telenet (Belgium) and KPN (The Netherlands) will deploy a 5G test network for this project. Only research participants of the project will be able to connect to it, using specifically provisioned SIM cards. On the prototype road vehicles (truck, cars and mobile crane), these SIM cards will only be used during specifically planned test days. Their users are always aware that they are participating to a specific test day, and hence this connectivity data is being recorded both at the mobile device and at the network side. Since the barge will not be a prototype, but belonging to the day-to-day operation of the shipping company, the mobile devices on the barge will automatically connect to the 5G network of pilot sites that they encounter on a day-to-day basis. The crew of that barge will be made aware of this.
Expected size of the data (if known)	The expected type and size of the data will be defined in T5.4 of WP5.
To whom might it be useful ('data utility')?	Participants of WP5 working on this network performance characterisation: Telenet, KPN, imec
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by IMEC, on the following storage infrastructure: central log server provided by IMEC. Access rights are arranged as follows: account will be provided by IMEC to the involved WP5 partners to get access to the central log server to upload and access the data. The data will be destroyed 60 months after the project, once no further audits on the results of the project can be expected.
Description of potential personal data, such as faces and license plates in recorded videos[2]	NA
Is the data sensitive according to the GDPR?	No, there is no sensitive data recorded

Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	- data classification - access control (account-based access, certificates, IP whitelisting, depending on data) - backups mandatory secure protocols and use of encrypted channels, e.g. VPN (to be defined in D2.3)
Usage of certified repositories for long term preservation and curation	Not applicable; the data will be deleted after 60 months after completion of the project.
Implemented anonymisation / pseudonymisation techniques.	Every device under test and each test session will be assigned with a unique ID. When sending to 3rd parties, these IDs are extra anonymized.
Privacy by design considerations	Not applicable (No personal/sensitive data)
Data breach protocol (how to handle at incidents)	There is an internal protocol available at IMEC, in which administrators are informed and they will give instructions on how to proceed.
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	Does not apply

2.4 Surveys

Surveys (WP3)	
Purpose of the data collection / generation	Capturing domain knowledge from actors in the teleoperation for transport and logistics value chain. This domain knowledge is needed to allow the expert researches regarding governance and business modelling to understand this specific application domain, so that they can create appropriate models and define appropriate parameter values in those models.
Relation to the objectives of the project.	This dataset is required to realize the following project objectives as defined in section 1.1 of part B of Annex 1 of the 5G-Blueprint Grant Agreement • BO1: 5G tele-operated market analysis • BO2: Commercial possibilities • BO3: Position the possible role of tele-operated transport based on 5G • BO4: Tele-operated transport based on 5G connectivity market adoption • RO1: Identify regulatory issues regarding the deployment of cross-border tele-operated transport based on 5G connectivity, and identify recommended actions.

Relevance and accordance with the 'data minimisation' principle in the envisaged use.	In order to capture this required domain knowledge, the surveyed actors will be asked questions in terms of characteristics of their day to day operation, their sector as a whole, the challenges they are faced with, the opportunities they see for teleoperation, etc. This dataset is considered to be the minimum dataset that is needed to support its purpose. This data will be captured on the individual organisation level so that it can be translated well into aggregated assumptions per actor type for adoption in the governance and business models. The individual organisation level input will however be kept on record to be able to verify the aggregated assumptions when needed.
Pilot site where the data will be captured (if applicable).	NA
Types and formats of the data	The data is collected in plain text.
Will existing data be re-used and how?	Data found in literature surveys will also be taken into account.
Origin of the data	The surveyed people will be employees of one of the following organisations: • Parties belonging to the project's group of beneficiaries • Parties belonging to the project's advisory board • Other relevant stakeholders which agreed to voluntarily be surveyed.
Expected size of the data (if known)	To be defined in D2.3.
To whom might it be useful ('data utility')?	The expert researches regarding governance and business modelling: imec, HZ, but also all other beneficiaries of the project.
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by …, on the following storage infrastructure: … (to be defined in D2.3). Access rights are arranged as follows: … (to be defined in D2.3). The data will be destroyed 60 months after the project, once no further audits on the results of the project can be expected.
Description of potential personal data, such as faces and license plates in recorded videos[2]	NA
Is the data sensitive according to the GDPR?	No. The surveys will be designed carefully to not result in the exposure of sensitive data.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	(to be defined in D2.3)
Usage of certified repositories for long term preservation and curation	(to be defined in D2.3)

Implemented anonymisation / pseudonymisation techniques.	(to be defined in D2.3)
Privacy by design considerations	(to be defined in D2.3)
Data breach protocol (how to handle at incidents)	(to be defined in D2.3)
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	(to be defined in D2.3)

2.5 Outreach & dissemination of results

Contact details newsletter registrations WP8	
Purpose of the data collection / generation	Allowing the project to distribute its perioded newsletter to interested individuals that registered for that newsletter, by capturing contact information as part of that registration process, and storing it for later use.
Relation to the objectives of the project.	This dataset is not required to realize any of the project objectives as defined in section 1.1 of part B of Annex 1 of the 5G-Blueprint Grant Agreement. However, it is required to realize the following objectives of WP8, as described in the corresponding work package description in part B of Annex 1 of the 5G-Blueprint Grant Agreement • To define and implement a comprehensive and effective set of dissemination and communication activities, creating awareness about project results and stimulating involvement of private and public stakeholders. • To facilitate exploitation of the project's outcome and actively promote the further development of innovative solutions based on the 5G-Blueprint outcome.
Relevance and accordance with the 'data minimisation' principle in the envisaged use.	Only the email address of the persons registering for the newsletter is requested for, and only as the consequence of an explicit opt-in of that user. Other typical personal data requested when subscribing to a newsletter are not asked to comply with the data minimisation principle (such as name, organisation or function).
Pilot site where the data will be captured (if applicable).	NA
Types and formats of the data	The data is collected in plain text.
Will existing data be re-used and how?	No.
Origin of the data	Visitors of the project website that decided to register for the project newsletter on the following URL: https://www.5gblueprint.eu/contact/ Note that on this project website the visitor has to possibility to get acquainted with the corresponding privacy policy:

	https://www.5gblueprint.eu/privacy-policy/
Expected size of the data (if known)	We collect minimal data from our site visitors
To whom might it be useful ('data utility')?	Data collected express the interest of subscribers to receive news from 5G-Blueprint project.
Details regarding storage of the data, including geographical details (inside/outside the EC), who has access rights to the data, and the time duration of storage	The data is stored by Martel, on the following storage infrastructure: GreenGeeks. Access rights are arranged as follows: role-based authorisation with administrator approval The data will be destroyed 60 months after the project, once no further audits on the results of the project can be expected.
Description of potential personal data, such as faces and license plates in recorded videos[2]	NA
Is the data sensitive according to the GDPR?	No, it is only an email address.
Data security provisions (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)	Prevention of unauthorised access: role-based authorisation on the GreenGeeks server. Data recovery: … Secure storage and transfer of sensitive data: NA.
Usage of certified repositories for long term preservation and curation	NA
Implemented anonymisation / pseudonymisation techniques.	NA
Privacy by design considerations	(to be defined in D2.3)
Data breach protocol (how to handle at incidents)	(to be defined in D2.3)
Sensitive non-personal data but with potential commercial impact, such as business cases, expert's personal opinions or 5G/network measurements: description, security provisions, data breach protocol, storage and processing.	(to be defined in D2.3)

3 FAIR PRINCIPLES APPLICABLE TO DATA

3.1 Introduction

According to the Template Horizon 2020 Data Management Plan which was used as the basis for this deliverable, this section should provide more details about the project approach to make the data Findable, Accessible, Interoperable and Re-usable (FAIR). It should answer the following questions:

- Findable
  - Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
  - What naming conventions do you follow?
  - Will search keywords be provided that optimize possibilities for re-use?
  - Do you provide clear version numbers?
  - What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.
- Accessible
  - Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.
  - Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.
  - How will the data be made accessible (e.g. by deposition in a repository)?
  - 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)?
  - Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.
  - Have you explored appropriate arrangements with the identified repository?
  - If there are restrictions on use, how will access be provided?
  - Is there a need for a data access committee?
  - Are there well described conditions for access (i.e. a machine readable license)?
  - How will the identity of the person accessing the data be ascertained?
- Interoperable
  - Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?
  - What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
  - Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?
  - In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
- Re-usable
  - How will the data be licensed to permit the widest re-use possible?
  - When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
  - Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
  - How long is it intended that the data remains re-usable?
  - Are data quality assurance processes described?

However, since due to the strategic nature of the expected results that will be developed within the project

the 5G-Blueprint consortium decided to opt-out the Pilot on Open Research Data (ORD) in Horizon 2020, no data will be made openly available. Hence the questions mentioned in the above bullet list are considered to be non-applicable.

Nevertheless, specific characteristics of the project's approach towards data and knowledge handling are related to the FAIR principles. Therefore these are introduced in the next subsections.

3.2 Making data findable, including provisions for metadata

Since the data collected or generated during this project solely will be used by the parties that were actively involved/engaged in its collection or generation, no specific measures are needed to make the data findable. Therefore no general project approach has been defined regarding metadata annotations to make the data discoverable, identifiable and locatable by means of a standard identification mechanism, nor regarding specific naming conventions, search keywords or version numbers (but as mentioned in section 2, almost all collected or generated data will be timestamped).

But instead of sharing data through the ORD, the project does commit to collaborating with external projects or initiatives as much as possible through knowledge transfer, and also, if feasible and at acceptable cost, through providing open access to the 5G-Blueprint pilot environment and deployed infrastructure (under certain agreed conditions and when not hindering the realization of the objectives of the 5G-Blueprint project).

3.3 Making data openly accessible

All published results arising from the research funded under 5G-Blueprint will be made openly accessible, in accordance with Horizon 2020's mandate on open access to all publications. These include all peer-reviewed publications and any resulting monographs, books, conference proceedings, and slides. Note that the possibility is not excluded that for specific scientific publications it will be decided to also include the underlying dataset in that publication. But that is to be decided on a case by case basis by the authors of the corresponding publication, and with agreement of all involved data owners, and with full compliance with all GDPR regulations.

Green open access will be favoured, where post-prints or publishers' PDFs of the research publications will be made available through one or more of the partners' institutional repositories, at no charge, and after an embargo period of between zero and six months has elapsed, depending on publisher policy.

Where green open access is not possible in a particularly desired target publication, the policy is to make a one-off payment, as per gold open access, to the publisher to ensure open access to the publication. This publication will then be accessible from the journal website, as well as being placed on the institutional repositories of one or more partners. Gold open access fees will come from the project budget. As open access is mandatory, publishers that do not allow green or gold open access will not be chosen for dissemination. Funding will be reserved in the project to support open access publication.

Public deliverables and presentations are stored on the "Library" section of the 5G-Blueprint website, reachable on the following URL: https://www.5gblueprint.eu/library/ .These same public deliverables are also automatically made available publicly on the Cordis website, which hosts all public project reports submitted by H2020 projects. The corresponding Cordis URL for this project is https://cordis.europa.eu/project/id/952189 . Any video material regarding the project or its results that is made publicly available will be shared through the project's YouTube channel, for which the URL is https://www.youtube.com/channel/UCv7n1u2SLeRH6DRJpfdGtrA .

3.4 Making data interoperable

To make the data as easy processable as possible, and interoperable across the boundaries of the different beneficiaries involved in the project, any collected or generated data will be formatted in a human-readable manner as much as possible, such as plain test, CSV, JSON, or XML. Since human-readable data storage can be considered as a security risk (higher impact in case of data breach), it will always be assed first if based on the Dataset characteristics human-readable formatting is appropriate or not.

3.5 Increase data re-use (through clarifying licenses)

To make sure that audits on the results of the project can be supported with the used data, all generated or collected data will be stored until 12 months after the end of the project. After that period they will be destroyed.

4 ALLOCATION OF RESOURCES

According to the Template Horizon 2020 Data Management Plan which was used as the basis for this deliverable, the following questions should be answered in this section:

- What are the costs for making data FAIR in your project?
- How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
- Who will be responsible for data management in your project?
- Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

As explained in section 3, since the 5G-Blueprint consortium decided to opt-out the Pilot on Open Research Data (ORD) in Horizon 2020, no data will be made openly available, and hence the collected and generated data is not made FAIR in this project. However, as described in section 3.3, all published results arising from the research funded under 5G-Blueprint will be made openly accessible as green open access or gold open access, and in case of public deliverables or presentations on the library section of the project website. The corresponding costs are covered by the project budget. Note that public deliverables are also made available and on the Cordis website of the European Commission, and video material regarding the project or its results are made publicly available on YouTube, but both these channels can be used without costs.

The overall responsibles for data management in the 5G-Blueprint project are the contributors to T2.3 “Quality insurance, risk management, ethical issues and Data Management Plan”: the Dutch Ministry of Infrastructure and Water Management (coordinator), and Martel Innovate (Project Management Office). However, the responsibility for the data management of the individual datasets (and for the resources for their preservation until 60 months after the end of the project) always lies at the party mentioned in the corresponding 5G-Blueprint Dataset Register record in section 2 as the one that takes care of the storage of the data. For each of those parties responsible for the data management of an individual dataset, a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project must be collected by the Coordinator. The corresponding details are captured in

Table 2: Overview of data management responsible per individual dataset

Beneficiary acronym	Contact details DPO or link to data protection policy for beneficiaries not required to appoint a DPO under the GDPR + contact details of beneficiary single point of contact	Dataset(s) name for which it is responsible
Seafar	Najmeh Masoudi, Najmeh.masoudi@seafar.eu	UC1
HAN	Karel Kural, Karel.Kural@han.nl	UC2
V-Tron	Rakshith Kusumakar, r.kusumakar@v-tron.eu	UC3
Roboauto	Oliver Held, oliver.held@roboauto.cz	UC4
Be-Mobile	Free Bruneel, free.bruneel@be-mobile.com	EF1
Locatienet	Tom Van de Ven, tom@traxpert.com	EF2
Swarco	Freek van der Valk, freek.vandervalk@swarco.com	EF3
imec	Ali Anwar, Ali.Anwar@imec.be	EF4
Roboauto	Oliver Held, oliver.held@roboauto.cz	EF5
Sentors	Sander Maas, sander.maas@sentors.nl	EF6

Be-Mobile	Free Bruneel, free.bruneel@be-mobile.com	EF7
R40	Wim Van den Broeck, wim@room40.technology	EF8
KPN	Matthijs Klepper, m.klepper@kpn.com	WP5 (network)
Imec	Pol Camps, Pol.Camps@imec.be	WP3 surveys
MAR	Maria Chiara Campodonico, mchiara.campodonico@martel-innovate.com	WP8 newsletter

5 DATA SECURITY

According to the Template Horizon 2020 Data Management Plan which was used as the basis for this deliverable, the following questions should be answered in this section:

- What provisions are in place for data security (including prevention of unauthorised access, data recovery as well as secure storage and transfer of sensitive data)?
- Is the data safely stored in certified repositories for long term preservation and curation?

Next to those questions, a third question needs to be answered in this section since Deliverable 1.2 “POPD – Requirement No. 2” refers to here: what are the anonymisation / pseudonymisation techniques that will be implemented?

The answer to these three questions can be different, depending on the specific dataset at hand. Therefore these three questions, and their corresponding answers, are integrated in the 5G-Blueprint Dataset Register presented in section 2.

6 ETHICAL ASPECTS

According to the Template Horizon 2020 Data Management Plan which was used as the basis for this deliverable, the following questions should be answered in this section:

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

Since the 5G-Blueprint consortium decided to opt-out the Pilot on Open Research Data (ORD) in Horizon 2020, no data will be made openly available, and hence no data sharing will be performed. Hence the questions mentioned in the above bullet list become non-applicable.

Note that even when opting out the ORD, ethical aspects have to be taken into account in the execution of the project, such as

- Procedure and templates for informed consent in compliance of the project ethical principles and data and privacy protection legislation and rules, based upon the GDPR.
- Description of potential ethical issues within the frame of the project such as Vulnerable Road Users interaction.

The way in which the project tackles these ethical aspects is described in detail in the project's D1.1 "H – Requirement No. 1".

When it comes to safeguarding the rights and freedoms of the research participants, the right to be forgotten is essential. For this one standard procedure is defined for the entire project: a research participant may always ask the DPO (of single point of contact if no DPO was not required under the GDPR) of the responsible for the dataset from which he/she wants to be removed to invoke this right to be forgotten. It is the responsibility of that person to take the needed appropriate steps to comply with this request, without any exception.

Note that in the 5G-Blueprint project no profiling will be involved, and that given the current content of the 5G-Blueprint Dataset Register presented in section 2, that no data protection impact assessment should be conducted, since none of the below situations that require the execution of a DPIA[6] are true for this project:

- a systematic and extensive evaluation of the personal aspects of an individual, including profiling;
- processing of sensitive data on a large scale;
- systematic monitoring of public areas on a large scale.

[6] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_en

7 OTHER ISSUES

According to the Template Horizon 2020 Data Management Plan which was used as the basis for this deliverable, the following questions should be answered in this section:

- Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

Consortium partners have not planned to apply any additional, specific data management procedures.

8 CONCLUSIONS

This document is the first version of the 5G-Blueprint Data Management Plan (DMP), defined at the end of M06. It is based on the Template Horizon 2020 Data Management Plan. According to the principles outlined in that template, and taking into account that at this moment in time the detailed requirements and technical architectures still are being analysed and defined, this first version of the DMP did not yet provide detailed answers on all questions listed in the DMP template. Instead, it was intended to constitute a living document in which information can be made available on a finer level of granularity through successive updates as the implementation of the project progresses and when significant changes occur, gaining more precision and substance during the lifespan of the project.

To validate the DMP framework presented in this deliverable, some first reflections regarding the type of datasets were worked out in a first draft version of the 5G-Blueprint Dataset Register. An important element significantly impacting the approach presented in this DMP is the fact that due to the strategic nature of the foreseen results part of the project, the partners decided to opt-out the Pilot on Open Research Data (ORD) in Horizon 2020. This was reflected in the presented details regarding FAIR data (Findable, Accessible, Interoperable, Re-usable), the allocation of resources (including DPO contact details where needed), data security, and ethical aspects.