D2.2

Data collection and evaluation procedures

www.5g-loginnov.eu

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No. 957400

Work Package	2
Task	2.2 Data collection and evaluation procedures
Authors	Djibrilla Amadou Kountche, Mandimby Ranaivo Rakotondravelona (AKKA), Pavlos Basaras (ICCS), Athanasios Koumparos (VFI), Luka Koršič, Janez Sterle, Rudolf Sušnik (ININ), Estíbaliz Loyo (VICOM) Johannes Chatzis (TEC4U)
Dissemination Level	Public
Status	FINAL
Due date	31/10/2021
Document Date	31/10/2021
Version Number	1.0

Quality Control

	Name	Organisation	Date
Editor	Mandimby Ranaivo Rakotondravelona	AKKA	08/10/2021
Peer review 1	Marco Gorini	CIRCLE	14/10/2021
Peer review 2	Dejan Šošter	TSLO	18/10/2021
Authorised by (Technical Coordinator)	Eusebiu Catana	ERTICO	21/10/2021
Authorised by (Quality Manager)	Mauro Dell'Amico	ICOOR	25/10/2021
Submitted by (Project Coordinator)	Eusebiu Catana	ERTICO	31/10/2021

Legal Disclaimer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
5GPPP	The 5G Public Private Partnership
API	Application Programming Interface
CAD/CAM	Connected Automated Driving / Connected and Automated Mobility
CAN	Controller Area Network
CCAM	Cooperative, Connected and Automated Mobility
DL	Downlink
DOI	Digital Object Identifier
EAMS	Enterprise Asset Management System
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
FOT	Field Operational Test
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
GPS	Global Position System
HTTPS	Hyper Text Transfer Protocol (HTTP) over Secure Socket Layer (SSL)
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LL	Living Lab
MNO	Mobile Network Operator
NSA	Non-Standalone
NFV-MANO	Network Function Virtualisation (NFV) - Management And Network Orchestration (MANO)
OBU	On-Board-Units
ORDP	Open Research Data Pilot
OSM	Open-source MANO
PCT	Piraeus Container Terminal
POPD	Protection Of Personal Data
PU	Public
QoE	Quality of Experience

REST	Representational state transfer
SME	Small and Medium Enterprise
TMT	Technical Management Team
TMS	Traffic Monitoring System
TRxP	Transmission Reception Point
UE	User Equipment
UL	Uplink
VIN	Vehicle Identification Number
VSaaS	Video Surveillance as a Service
WP	Work Package

EXECUTIVE SUMMARY

The evaluation phase of 5G-LOGINNOV as defined in D1.4 – "Initial specification of evaluation and KPIs" requires the collection of data from the Living Labs trials. This deliverable provides details on the data collection architecture and the tools to be implemented and deployed to this purpose. In particular, it describes the central data collection tool which gathers all the evaluation data and makes them available to the evaluation team (tasks 3.5 and 3.6 and LL leaders).

The first chapter introduces 5G-LOGINNOV with general information on the current deliverable such as its links with the other tasks and deliverables.

Chapter 2 focuses on the data collection architecture by highlighting the requirements that drove its design as well as its specification. They have been categorised as common, evaluation-specific, Living Lab-specific and central data collection tool-specific requirements.

Chapter 3 addresses the evaluation data management and describes the approach, the metadata and the data that will be handed over to the central data collection tool. This chapter is of high importance for the evaluation tasks as it provides the description of all the evaluation data that will be at the evaluation team disposal.

Chapter 4 gives an overview of the Living Labs local data collection architecture and tools that will interface with the central data collection tool.

Chapter 5 provides details on the central data collection tool, its components as well as how it is implemented and deployed. The central data collection tool is responsible for ingesting the evaluation data from the Living Labs, indexing and storing them while making them available through an API or a user interface. It also allows the publication of some data to an open repository in the frame of ORDP.

Chapter 6 concludes the deliverable with insights on future work.

1 INTRODUCTION

1.1 Project introduction

5G-LOGINNOV will focus on seven 5G-PPP thematics and support to the emergence of a European offer for new 5G core technologies in 11 clusters of use cases.

5G-LOGINNOV's main aim is to design an innovative framework addressing integration and validation of Connected Automated Driving/Mobility (CAD/CAM) technologies related to the industry 4.0 and ports domains by creating new opportunities for LOGistics value chain INNOVation.

5G-LOGINNOV is supported by 5G technological blocks, including new generation of 5G terminals notably for future Connected and Automated Mobility, new types of Industrial Internet of Things 5G devices, data analytics, next generation traffic management and emerging 5G network architectures, for city ports to handle upcoming and future capacity, traffic, efficiency and environmental challenges.

5G-LOGINNOV will deploy and trail 11 clusters of Use cases targeting beyond TRL7 including a GREEN TRUCK INITIATIVE using CAD/CAM & automatic trucks platooning based on 5G technological blocks.

Thanks to the new advanced capabilities of 5G relating to wireless connectivity and Core Network agility, 5G-LOGINNOV ports will not only significantly optimize their operations but also minimize their environmental footprint to the city and the disturbance to the local population.

5G-LOGINNOV will be a catalyst for market opportunities build on 5G Core Technologies in the Logistics and Port operations domains, thus being a pillar of economic development and business innovation and promoting local innovative high-tech SMEs and Start-Ups. 5G-LOGINNOV will open SMEs' and Start-Ups' doors to these new markets using its three Living Labs as facilitators and ambassadors for innovation in future European ports.

5G-LOGINNOV's promising innovations are key for the major deep-sea European ports in view of the mega-vessel era (Hamburg, Athens), and are also relevant for medium-sized ports with limited investment funds (Koper) for 5G.

1.2 Purpose of the deliverable

The main purpose of this deliverable is to describe the tools that will be implemented and deployed for evaluation data collection. To this end, this document first provides an overview of the requirements that drove the design of the data collection architecture and tools. Then, from the requirements, the architecture is specified, and the evaluation data management approach is defined. In particular, metadata and descriptions of the data that will be handed over to the central data collection tool are available in this document. Finally, technical details regarding the central data collection tool are also provided.

This deliverable gives the reader insights on the evaluation data collection tools, specifically the central one, from the design to the implementation. Moreover, this document can be used by the Living Labs to access the necessary technical information in order to interface with the central data collection tool. Additionally, this deliverable will support the evaluation tasks by providing metadata and evaluation data description to the evaluation team.

1.3 Intended audience

This deliverable is PUBLIC intended for the following audiences:

- 5G-LOGINNOV partners must use the deliverable as recorded agreement, reference and guideline throughout the development and deployment of the innovations.
- The EU and related reviewers can use the deliverable to gain insight into how the development and deployment work result in the delivery of the 5G-LOGINNOV innovations.
- Any reader can use the deliverable to gain insight in how these kinds of innovations are tracked and realized in 5G-LOGINNOV.

1.4 Structure of the deliverable and relation with other work packages/deliverables

This deliverable is structured as follows:

- Chapter 1 introduces 5G-LOGINNOV with general information on the current deliverable.
- Chapter 2 focuses on the data collection architecture by highlighting the requirements that drove its design as well as its specification.
- Chapter 3 addresses the evaluation data management describing the approach, the metadata and the data that will be handed over to the central data collection tool.
- Chapter 4 gives an overview of the Living Labs local data collection tools.
- Chapter 5 provides details on the central data collection tool, its components as well as how it is implemented and deployed.
- Chapter 6 concludes the deliverable with insights on future work.

This deliverable takes as input the data management plan defined in D6.4 – "Data management plan" as well as the requirements defined in D1.4 – "Initial specification of evaluation and KPIs" and D1.5 – "Data and cyber protection policies". Additionally, the Living Labs storyboards described in D3.1 – "Trial methodology, planning and coordination" have been considered to design the metadata. The output of this deliverable will serve the T2.3, T2.4 and T2.5 tasks which are about the development and deployment of the Living Labs especially regarding the interfacing with the central data collection tool. Finally, this deliverable will support evaluation tasks T3.5 and T3.6 by providing details on the evaluation data collection tools as well as a description of the data.

2 DESIGN OF THE DATA COLLECTION ARCHITECTURE

2.1 Requirements

This section provides the identified requirements related to data collection. They have driven the specification of the data collection architecture (see 2.2). They result from a requirement elicitation process started in T1.4 relying on a shared Excel file listing all identified requirements and discussions with the main stakeholders, namely the data manager, the data producers including the Living Labs and the data users including the evaluation team.

For each requirement, the following details are given as follows:

- ID: unique identifier providing also the category (see below)
- Topic: the general topic in which the requirement applies
- Description: the requirement statement. The used modal verb indicates the priority of the requirement based on the MoSCow method[1].

The requirements are grouped in 4 categories as described in the following subsections:

- Common requirements
- Evaluation requirements
- Living Labs requirements
- Central data collection tool requirements

2.1.1 Common requirements

The common requirements consist of general requirements that are applicable to all from the data producers to the data users through the tools used for data management. They are listed in Table 1. Their IDs are prefixed with CR.

ID	Topic	Description
CR-01	Data management	5G-LOGINNOV must apply the FAIR principles as stated in D6.4 (Data Management Plan)
CR-02	Data management	5G-LOGINNOV must apply the ORDP principles as stated in D6.4
CR-03	Data management	All the provided data must be clearly identified with metadata and timestamped
CR-04	Data privacy	GDPR requirements stated in D6.4 and WP7 (Ethics) deliverables must be followed
CR-05	Data privacy	All collected data must be anonymised before storage
CR-06	Data quality	The data producers must have the necessary tools to check data quality
CR-07	Data quality	The data producers must provide the requested data in compliance with evaluation requirements (2.1.2)

[1] https://www.agilebusiness.org/page/ProjectFramework_10_MoSCoWPrioritisation

CR-08	Data quality	The data producers must provide descriptions of the evaluation data to be collected
CR-09	Data quality	The data producers must provide the requested data in compliance with the provided data descriptions
CR-10	Data security	The data must be exchanged using a secured transport layer
CR-11	Data storage	The data collection tools must provide a secured storage and backup to avoid data loss
CR-12	Synchronization	Universal Time Coordinated (UTC) must be used as the single time reference
CR-13	Synchronization	All applications providing logging, and all log data provided, shall be time synchronised

Table 1 Common requirements

2.1.2 Evaluation requirements

The evaluation requirements include the requirements on the data collection to ensure qualitative evaluation tasks. The following requirements are based on D1.4 and anticipated discussion with WP3 in which the evaluation tasks will be performed (T3.5 and T3.6). IDs are prefixed with ER and are listed in Table 2.

ID	Topic	Description
ER-01	KPI	The evaluation team must provide the KPIs necessary for the evaluation tasks
ER-02	KPI	The data producers must provide the data necessary for KPIs calculation
ER-03	Data collection	The collected data must be documented with precise metadata
ER-04	Data collection	The collected data shall be raw or processed data
ER-05	Data access	The tool that will store the evaluation data shall accept scripted and SQL-like requests

Table 2 Evaluation requirements

2.1.3 Living Labs requirements

The Living Labs requirements correspond to the requirements directed to the LLs in a global way, but may include specific requirements applicable to a single LL. They address the integration of the LLs with the data collection process as well as the quality of the locally collected data. They are provided in Table 3 with LR as ID prefix.

ID	Topic	Description
LR-01	Data management	Each LL can define a common data model
LR-02	Data management	Each LL must provide the data model of the collected evaluation data
LR-03	Data quality	Each LL must document the collected evaluation data with relevant metadata

LR-04	Data ingestion	Each LL must set up an interface with the central data collection tool
LR-05	Athens LL	Telemetry data from the fleet of 5G connected trucks must be aggregated and processed for fuel consumption, average travel distance and CO2 emissions calculation
LR-06	Athens LL	Machine learning model must be provided for human presence detection
LR-07	Athens LL	Computer vision model must be provided for container seal detection
LR-08	Koper LL	Annotated data must be provided for the evaluation of the model accuracy
LR-09	Koper LL	Images from all the faces of containers must be provided with a determined quality
LR-10	Koper LL	Specifications of the capturing sensors must be provided
LR-11	Koper LL	Images from a high view perspective must be provided with a determined quality
LR-12	Koper LL	Lighting specifications must be provided

Table 3 Living Labs requirements

2.1.4 Central data collection tool requirements

The central data collection tool requirements are directed towards the central tool that will be specified and developed to ingest, store and provide the collected evaluation data (see chapter 5). The requirements include functional and non-functional ones. They are listed in Table 4 with CTR as ID prefix (central tool requirements).

ID	Topic	Description
CTR-01	Data access	The central tool must provide an interface to download data
CTR-02	Data access	The central tool must provide authentication and authorization features
CTR-03	Data access	The central tool must offer a user interface for the data users
CTR-04	Data ingestion	The central tool must only accept raw or processed data compliant with the agreed data descriptions
CTR-05	Data quality	The central tool must perform a quality check of the ingested data
CTR-06	Data storage	The central tool must persist data
CTR-07	Data storage	The data stored on the central tool must be available until the end of the project
CTR-08	Data storage	The central data collection tool must provide enough storage space for all the provided data

Table 4 Central data collection tool requirements

2.2 Specification

2.2.1 High-level architecture

FOT-Net—a European Commission-funded Support Action regarding Field Operational Tests (FOTs) [1], describes an approach called the "space mission" approach in "*which as much data as possible are collected, because the FOT provides a unique opportunity (and funding) to collect data which may be hard to collect later on. This approach gives a rich dataset, that enhances the probability that the data will be re-used in future projects. However, before starting data collection, it is recommended to develop a plan on how to store the data and how to make them available for later analysis or analysis by others. This plan should specify detailed data dictionaries, open software formats, rules for data access and other relevant information as meta-data*". Figure 1 describes the high-level architecture devised for 5G-LOGINNOV in order to collect a rich dataset tailored to the project evaluation needs.

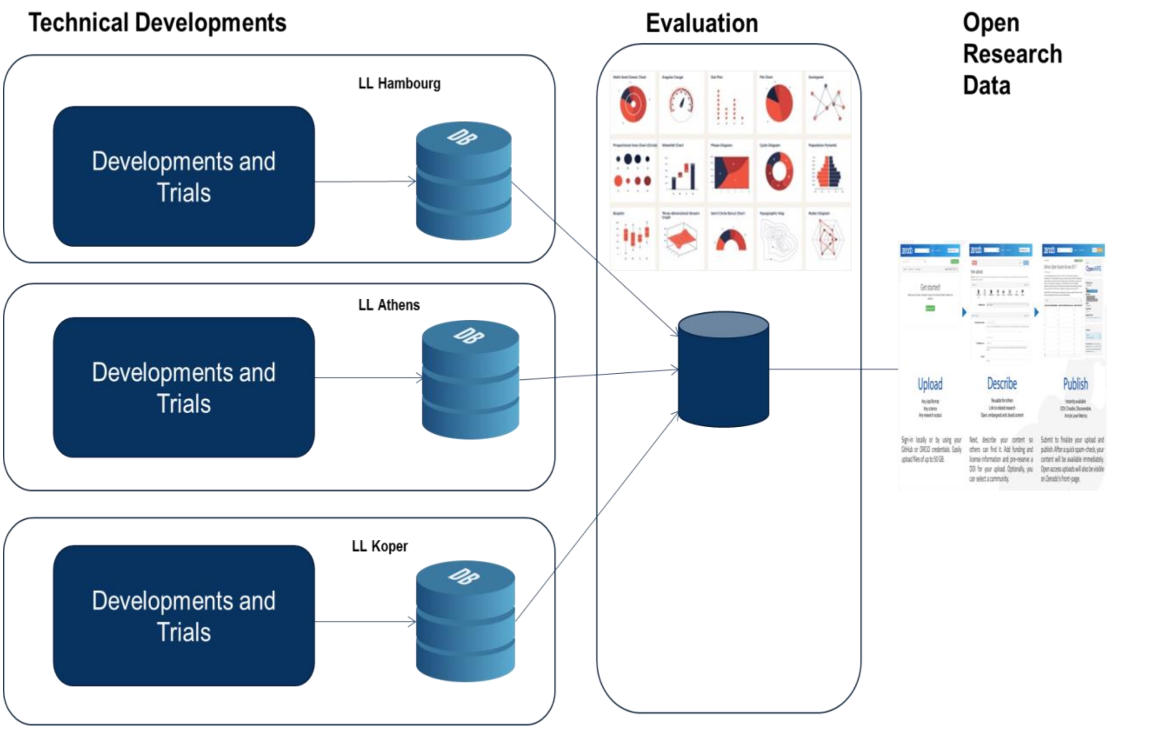

Figure 1: Data management in 5G-LOGINNOV

2.2.2 Interfaces

The project's Data Management Plan (D6.4) [2] has provided an initial high-level architecture for data management as illustrated in Figure 1. It shows that each of the three Living Labs generates data during the use case operations and trials. The data may be raw or processed (e.g. KPIs). Some of the generated data are collected to a central data collection tool for evaluation purposes. It is up to the Living Labs to decide what data they are willing to provide to support the evaluation tasks.

The central data collection will interface with each of the Living Labs to collect evaluation data. As such, it makes project-wide evaluation data available to the evaluation team. Additionally, the central data collection tool will help in publishing selected data under ORDP (Open Research Data Pilot). The deliverable D1.4 – Initial specification of evaluation and KPIs [3] provides input for data collection by providing a list of KPIs and the data necessary to calculate them. It provides also common requirements regarding evaluation data that will be considered.

The data collection process will consider the data handling requirements and cybersecurity policies described in the deliverable D1.5 – Data and cyber protection policies [4].

2.2.2.1 Interfaces with Living Labs data collection tools

The Living Labs will use a unique interface provided by a data aggregator to send data to the central data collection tool. This interface is a standardised protocol defined at length in the next section. These data will undergo some pre-processing defined below (as described in FOT-NET [1]:

1. ***Assessing and quantifying missing data*** *(e.g. percentage of data actually collected compared to the potential total amount of data which it was possible to collect).*
2. ***Ensuring that data values are reasonable and units of measure are correct*** *(e.g., a mean speed value of 6 may be unreasonable unless speed was actually recorded in m/s instead of km/h).*
3. ***Checking that the data dynamic over time is appropriate for each kind of measure*** *(e.g., if the minimum speed and the maximum speed of a journey are the same, then the data may not have been correctly sampled).*
4. ***Guaranteeing that measures features satisfy the requirements for the specific data analysis*** *(e.g., in order to calculate a reliable value of the standard deviation of lane offset, the lane offset measure should be at least 10s long; additionally, this time length may depend on the sampling rate)"*

The first three sub-steps, illustrated in Figure 2, refer to general quality checks; thus, if any of these fails, data analysis cannot proceed. If a failure is encountered, it should then be reported to the Living Labs by the Evaluation data aggregator so that the possible technical errors can be tracked down and solved. These steps and the error reporting between the central data collection tools and the Living Labs will be implemented in chapter 5. However, the last sub-step is different, and is related to the specific analysis or to a specific KPI to be used in the subsequent data analysis. As a consequence, if step 4 fails, it may not be due to a technical issue that needs to be solved, but to intrinsic limitations in the collected data [1]. To ensure the success of the last step, several rounds of discussions have been held between the Living Labs, the Evaluation Team and the Data Manager to derive the data described in section 3.4. These discussions will continue till the end of the project.

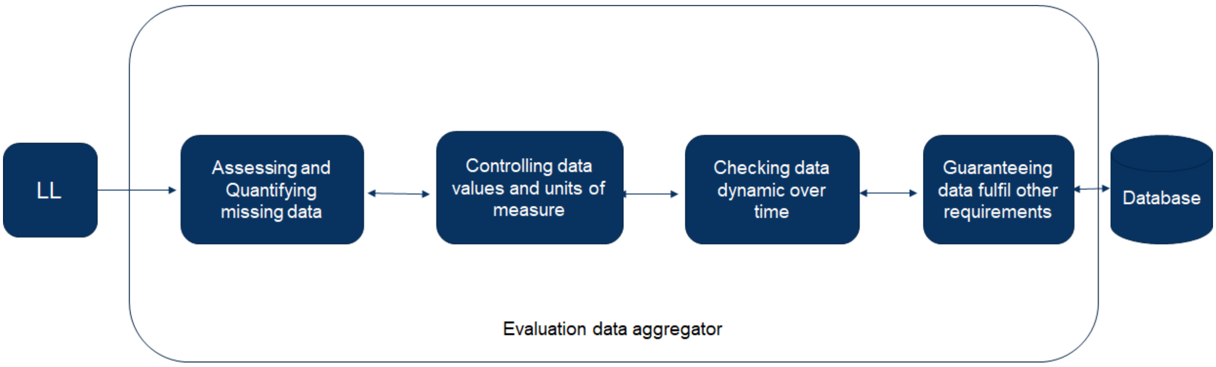

Figure 2: Evaluation data aggregator interface

The implementation of this interface will be described in section 5.2.1.

2.2.2.2 Interfaces with the evaluation team

The Evaluation Team comprises the member of Task 3.5 – Evaluation of operation optimization and Task 3.6 – Evaluation of social, economic and environmental impacts and the Living Lab leaders. This team is responsible for conducting the evaluation of the project based on the methodology defined in the Deliverable D1.4 – Initial specification of evaluation and KPIs [3]. This evaluation is based on the data collected by the data collection tools. Therefore, the evaluation team interacts with the tools by the following steps:

- Interact through a user interface which can be an application or an API
- Evaluate the project by analysing the data and computing the KPIs

- Provide the results of these evaluations to the central data collection tool

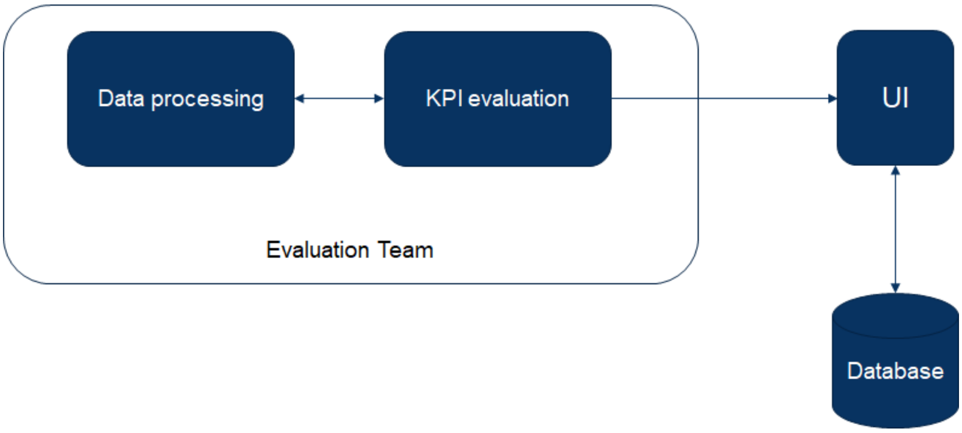

Figure 3: Evaluation team interface

The implementation of this interface will be described in section 5.2.3.

3 EVALUATION DATA MANAGEMENT

3.1 Principles

The data production occurs in the Living Labs during the execution of the storyboards [5]. However, each storyboard is executed in a different environment and setup and requires the implication of several partners. Therefore, in order to prevent data loss and ensure the coherence of the collected data, a coordination is mandatory between the partners of a Living Lab. Each partner ensures the acquisition of certain data and applies internal data processing. The data acquisition step can be an automatic or a manual process. After the acquisition, the data are stored in the partner's own storage before being transferred to the Living Lab's central data storage. From the Living Labs data storages, the evaluation data is transferred to the central data collections tool with the necessary metadata (see 3.3). This process is illustrated in Figure 4.

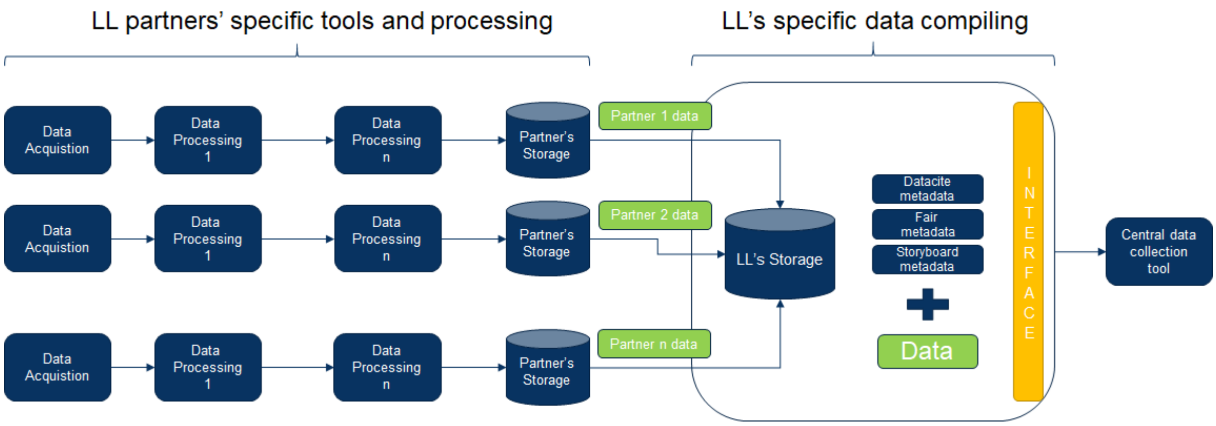

Figure 4: Living Labs local data acquisition, processing and storage

3.2 Approach

As described in the previous section, several types of data will be produced during the execution of the storyboards. The evaluation data are enriched with a common metadata format to satisfy FAIR (Findability, Accessibility, Interoperability, and Reuse) requirements [6] (CR-1, CR-2 and CR-3 requirements). Hence, the evaluation data are composed of two parts: a common metadata header and the provided Living Lab data.

The metadata are originated from different recommendations such as DataCite [7], FAIR [6] and 5G-LOGINNOV storyboards as described in more details in section 3.3.

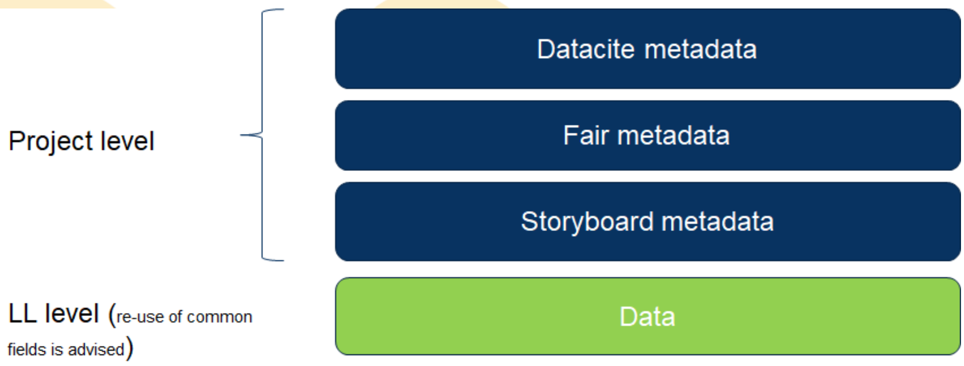

Figure 5: 5G-LOGINNOV evaluation data

3.3 Metadata

As shown in Figure 5, 5G-LOGINNOV's metadata are a combination of several properties taken from the DataCite metadata schema [7], FAIR metadata and storyboard-related metadata. They are described in the following subsections.

Each category of metadata will be detailed in a table with the following columns:

- Property: the name and identifier of the metadata property
- Sub-property: the name and identifier of the metadata sub-property, if any
- Required/Optional: indicates whether it is mandatory to provide the property or not
- Description: indicates what data are expected in the field

3.3.1 DataCite metadata

DataCite is a non-profit organisation acting as a persistent identifiers (DOIs) provider. DataCite proposes an extensive metadata schema [7] that has been adopted by other persistent identifiers providers helping to make research data FAIR. A subset of the DataCite metadata fields will be used for 5G-LOGINNOV evaluation data as shown in Table 5.

Property	Sub-properties	Required/ Optional	Description
Identifier		Required	Unique identifier of the dataset.
	identifierType	Required	Type of the identifier such as DOI.
Creator		Required	People involved in data production. To be repeated in case of multiple creators.
	creatorName	Required	Full name of the creator.
	nameType	Optional	Organizational/Personal.
	affiliation	Optional	Organizational or institutional affiliation of the creator.
Title		Required	Title of the dataset.
Publisher		Required	Name of the entity that holds, archives, publishes the dataset.
PublicationYear		Required	Year when the data was or will be made publicly available.
ResourceType		Required	Description of the resource. E.g. Dataset/Census data
	reourceType General	Required	General type of the resource. E.g. Dataset. Full list available in [7].
Subject		Optional	Subject, keyword, classification code, or key phrase describing the resource.
Contributor		Optional	People who contributed toto data collection, management and publication. To be repeated in case of multiple contributors.

	contributorType	Required	Type of contributor. E.g. data collector, data manager … Full list available in [7].
	contributorName	Required	Full name of the contributor.
	nameType	Optional	Organizational/Personal.
	affiliation	Optional	Organizational or institutional affiliation of the contributor.
Date		Optional	Different dates relevant to the work.
	dateType	Required	Type of date. E.g. Created, submitted. Full list in [7].
Language		Optional	Primary language of the resource. E.g. en, de …
RelatedIdentifier		Optional	Identifiers of related resources.
	relatedIdentifierType	Required	Type of the RelatedIdentifier. E.g. DOI. Full list in [7].
	relationType	Required	Description of the relationship with the related resource. E.g. IsDerivedFrom, IsPartOf … Full list in [7].
Size		Optional	Size of the resource. E.g. 6 MB
Format		Optional	Technical format of the resource. E.g. text/xml
Version		Optional	Version number of the resource.
Rights		Optional	Any rights information for the resource. E.g. license. Additionally, in D1.5 [4], we have 4 levels of access rights: open, restricted, closed and embargoed
Description		Optional	All additional information that does not fit in any of the other categories.
	descriptionType	Required	Type of description. E.g. Abstract, methods, technical info … Full list in [7].
GeoLocation		Optional	Spatial region or named place where the data was gathered.
FundingReference		Optional	Information about financial support (funding).
	funderName	Required	Name of the funding provider.

Table 5: DataCite metadata

3.3.2 FAIR metadata

FAIR metadata consists of all metadata required to make data FAIR [6]. FAIR metadata and DataCite metadata serve the same objective, hence there is some overlapping between them as can be seen below. FAIR metadata can be divided into three categories[2]:

- Administrative metadata: they correspond to management data about the resource such as the owner, the collaborators, the funding etc. They are already covered by the DataCite metadata.
- Descriptive metadata: they consist in the data describing the resource allowing its discovery and identification such as the unique identifier, the title, the abstract etc. They are also covered by the DataCite metadata.
- Structural metadata: they provide information on how the resource has been constituted and how the resource is internally organized. The first part can be covered by the DataCite metadata in the Description field, where information such as methods or any technical info can be provided. The second part describes the content of the resource. Hence, it corresponds to the data model that describes the referred dataset providing information on the data fields, the data types, the units etc.

To sum up, for each dataset, in addition to the DataCite metadata, the relevant data model must be provided as part of the FAIR metadata. The description of the collected evaluation data is provided in section 3.4.

3.3.3 Storyboard metadata

Storyboards as defined in D3.1 – Trial methodology, planning and coordination [5] provide the relevant information to set up and perform the use cases of the demonstrations. As such, they describe the context in which the evaluation data have been collected. Hence, the description of the relevant storyboard will be provided as valuable metadata for each dataset.

Table 6 describes the properties and sub-properties of the storyboard metadata based on the storyboard template provided in D3.1 [5].

Property	Sub- properties	Required/ Optional	Description
StoryboardID		Required	Numeric identifier for each Living Lab for the storyboard
StoryboardTitle		Required	Name of the storyboard
UC		Required	Name of the concerned use case
IterationID		Required	Numeric identifier of the iteration. Not in the storyboard template but valuable information.
KPI		Required	List of the relevant KPIs for the storyboard
BaselineData		Required	Description of the approach to collect baseline data (Level KPI)
OperationalData		Required	Description of the approach to collect operational data (Level KPI)
EvaluationData		Required	Description of the approach to provide data for evaluation (Level KPI)

[2] https://www.howtofair.dk/how-to-fair/metadata/

Setup	Organizational	Required	The organizational 'setup': e.g. Vehicles, infrastructure, participants etc.
	Technical	Required	The technical setup to process the storyboard with regards to WP2 architecture and the overall technical bracket related to 5G technologies
	Other	Optional	Optional information about 'story' and 'setup' e.g. diagrams, maps, pictures etc.

Table 6: Storyboard metadata

3.4 5G-LOGINNOV data description

This section provides information on the structure of the collected evaluation data that will be handed over by the Living Labs to the central data collection tool. The collected data are identified in the storyboards provided in D3.1 [5] in the "evaluation data" section. As stated in 3.3.2, the data descriptions serve as FAIR metadata for the internal descriptive part. Moreover, the data descriptions are necessary for validating the data ingested in the central data collection tool and will serve to design the mapping templates in the implementation phase (see chapter 5).

For each dataset, the following information is provided:

- Field: the name of the data field
- Type: the type of the field. E.g. string, integer …
- Unit: the unit of the value of the field
- Required/Optional: indicates whether it is mandatory to provide the field or not
- Description: short explanation of what is the data field about

3.4.1 Athens LL data description

According to the storyboards, the Athens LL will hand over the following data to the central tool.

3.4.1.1 Technical/5G KPI data

Model inference time (A-KPI9 & A-KPI14)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Inference time	Float	Seconds	Required	The time required to process the input of video stream(s) and infer the presence/absence of people (UC4) OR to infer the presence/absence of container seals (UC5)
Number of detections	Int	-	Required	Number of people detected (UC4) or number of seals detected (UC5)
Type	String	-	Required	Seal or Person
Timestamp	String	-	Required	
Frame id	Int	-	Optional	The video frame that the inference was based on.

Model accuracy/reliability (A-KPI10 & A-KPI15)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
True Positives	Int	-	Required	Correct inference of actual presence (human or seal presence, UC4 and UC5)
True Negatives	Int	-	Required	Correct inference of actual absence (human or seal absence, UC4 and UC5)
False Positives	Int	-	Required	False inference of presence (human or seal presence, UC4 and UC5)
False Negatives	Int	-	Required	False inference of absence (human or seal absence, UC4 and UC5)
Type	String	-	Required	To distinguish between human presence detection (UC4) and container seal detection (UC5)
Start Date	Date	-	Required	
End Date	Date	-	Required	

Time needed the device to open a network connection (A-KPI7)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Timestamp	String	-	Required	Timestamp of modem network attach
Duration	Int	Milliseconds	Required	Duration of the measurement, i.e., the amount of time between modem network attach and packet sends.
Device Id	String	-	Required	The IoT device id

User experienced data rate (A-KPI11)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Data rate	Float	Mbps	Required	Experienced uplink data rate from the 5G-IoT device to PCT management platform [UC4, UC5]
Timestamp	String	-	Required	
Duration	Int	Seconds/ minutes	Required	Duration of the measurement, e.g., the amount of time the live video transmission service (or other traffic) is active
Device Id	String	-	Required	The IoT device id that triggers the live video session

Deployment Time (A-KPI12)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Duration	Float	Seconds	Required	MANO service orchestration/deployment time of computer vision analytics to 5G-IoT devices [UC4 – Human presence detection, UC5 – Container seal detection].
Timestamp	String	-	Required	
Device Id	String	-	Required	The IoT device that the service instantiation was executed
Type	String	-	Optional	Human presence or Container seal service

Area traffic capacity (A-KPI21)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Area traffic capacity	Float	Mbps	Required	The total traffic throughput served per geographic area via a 5G basestation.
Timestamp	String	-	Required	
Basestation Id	String	-	Required	The 5G basestation Id
Duration	Int	Seconds	Optional	The duration of the experiment

Bandwidth (A-KPI22)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Bandwidth	Float	Mbps	Required	Maximum TCP/IP uplink and downlink bandwidth measured from the end user device on 5G RAN to the reference server located in 5G core.
Timestamp	String	-	Required	
Test Type	String	-	Required	Uplink or downlink measurement
Device Type	String	-	Required	Stationary, e.g., 5G IoT, or mobile, e.g., 5G Truck.
Target	IP	-	Required	IP address of test end point (any public or private IPv4 address)
Basestation Id	String	-	Required	The 5G base station Id

Connection density (A-KPI23)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Number of connected devices	Int	Per km2	Required	The total number of connected and/or accessible devices per unit area.
Timestamp	String	-	Required	
Basestation Id	String	-	Required	The 5G base station Id.

Reliability (A-KPI24)				
Field	**Type**	**Unit**	**Required /Optional**	**Description**
Reliability	Float	%	Required	The percentage (%) of the amount of sent network layer packets successfully delivered to a given system node (including the User Equipment) within the time constraint required by the targeted service, divided by the total number of sent network layer packets.
Target	IP	-	Required	IP address of test end point (any public or private IPv4 address)
Timestamp	String	-	Required	
Basestation Id	String	-	Required	The 5G base station Id.

End-to-end latency (A-KPI25)				
Field	**Type**	**Unit**	**Required/ Optional**	**Description**
e2e-latency	Int	Milliseconds	Required	Measured round trip time (RTT) from the moment the IP ICMP Echo Request packet leaves the source host until the IP ICMP Echo Reply is received from the destination host.
Device Type	String	-	Required	Stationary, e.g., 5G IoT, or mobile, e.g., 5G Truck.
Target	IP	-	Required	IP address of test end point (any public or private IPv4 address)
Timestamp	String	-	Required	
Basestation Id	String	-	Required	The 5G base station id

One-way latency (A-KPI26)				
Field	**Type**	**Unit**	**Required/ Optional**	**Description**
one-way latency	Float		Required	The one-way latency is the total time that is required for a packet to be generated at the communication unit at the transmitter's side, until it is received at the communication unit at the receiver's side.
Device Type	String	-	Required	Stationary, e.g., 5G IoT, or mobile, e.g., 5G Truck.
Target	IP	-	Required	IP address of test end point (any public or private IPv4 address)
Timestamp	String	-	Required	
Basestation Id	String	-	Required	The 5G base station id

3.4.1.2 Logistics KPI data

CO2 emissions (A-KPI1)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Fuel consumption	Float	Litre	Required	Fuel consumption of yard trucks, directly available from CAN-Bus (average daily weekly or monthly values).
Truck travel distance	Float	km	Required	Travel distance of yard trucks, directly available from CAN-BUS (average daily weekly or monthly values).
Type	String	-	Required	4G or 5G truck to differentiate between baseline and operational data.
Start Date	String	-	Required	Will be used for both, truck travel distance and fuel consumption records (daily, weekly or monthly).
End Date	String	-	Required	Will be used for both, truck travel distance and fuel consumption records (daily, weekly or monthly).

Fuel consumption (A-KPI2)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Fuel consumption	Float	Litre	Required	Average fuel consumption of trucks, directly available from CAN-Bus data (average daily weekly or monthly values).
Type	String	-	Required	4G or 5G truck to differentiate between baseline and operational data.
Start date	Date	-	Required	Start date measurement of average fuel consumption
End date	Date	-	Required	End date measurement of average fuel consumption

Truck travel distance (A-KPI3)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Truck travel distance	Float	km	Required	Truck travel distance, directly available from CAN-BUS data (average daily weekly or monthly values).
Type	String	-	Required	4G or 5G truck to differentiate between baseline and operational data.
Start date	Date	-	Required	Start date measurement of average truck travel distance
End date	Date	-	Required	End date measurement of average truck travel distance

Assets idling (A-KPI4)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Duration	Int		Required	Time duration where 4/5G yard trucks remain idle, not participating in container jobs.
Type	String	-	Required	4G or 5G truck to differentiate between baseline and operational data.
Start Date	Date	-	Required	Start date measurement of average idle time
End Date	Date	-	Required	End date measurement of average idle time

Percent of empty containers runs (A-KPI5)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
TruckId	String	-	Required	External truck Id
Location	String	-	Optional	Indicates the location of the truck inside or outside the port.
Entry Timestamp	String	-	Required	Will be used for truck arrival at the port
Duration	int	Seconds	Required	Duration of stay
OnLoadArea	Boolean	-	Required	Based on location we identify if the truck is loaded or not.

Mean time of container job (A-KPI6)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
TruckId	String		Required	External truck Id
Location	String	-	Required	Indicates the location of the truck inside or outside the port.
Type	String	-	Required	To distinguish between historical and trial data
Entry Timestamp	String	-	Required	Will be used for truck arrival at the port
Duration	int	Minutes	Required	Duration of stay

Human resource optimization (A-KPI8)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Person Hours	Float	Hours	Required	Person hours (or deviation of hours) spent for security or surveillance patrols before and after UC4; the person hours (or deviation of hours) spent for manual seal check before and after the activation of UC5.
Type	String	-	Required	To differentiate between UC4 and UC5 trial data from baseline/historical data, or the deviation in the person hours.
Start date	String	-	Required	Start date measurement of average asset idle time.
End date	String	-	Required	End date measurement of average asset idle time.

Vessel operation completion time (A-KPI13)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Vessel completion time	Float	Hours	Required	Vessel stay time (or deviation of stay time) at PCT premises (before and after UC5).
Type	String	-	Required	To differentiate between historical and trial data.
Start date	String	-	Required	Start date measurement of average vessel stay time.
End date	String	-	Required	End date measurement of average vessel stay idle time.

Parts in stock (A-KPI16)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Part type	String	-	Required	The part type that the predictive maintenance algorithm is focused on (different part types will be involved).
Consumption	Int	-	Required	Number of parts consumed/used of each part type evaluated (average).
Type	String	-	Required	To distinguish historical logged data from trial data.
Start date	String	-	Required	Start date measurement of average part in stock data.
End date	String	-	Required	End date measurement of average parts in stock data.

Vehicle breakdown (A-KPI17)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Breakdown frequency	Int	-	Required	Number of yard truck breakdown events.
Type	String	-	Required	To distinguish historical logged data from trial data.
Start date	String	-	Required	Start date measurement of average break down frequency.
End date	String	-	Required	End date measurement of average parts in stock data.

Vehicle under maintenance (A-KPI18)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Truck downtime	Float	Hours	Required	Deviation in the downtime of yard trucks for repairs.
Type	String	-	Required	To distinguish historical logged data from trial data.
Start date	String	-	Required	Start date measurement of average truck downtime.
End date	String	-	Required	End date measurement of average truck downtime.

Vehicles unexpected breakdown (A-KPI19)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Unexpected breakdown frequency	Int	-	Required	Number of unexpected vehicle breakdowns.
Type	String	-	Required	To distinguish historical logged data from trial data.
Start date	String	-	Required	Start date measurements of unexpected vehicle breakdowns.
End date	String	-	Required	End date measurements of unexpected vehicle breakdowns.

Maintenance costs of vehicle (A-KPI20)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Maintenanc e cost	Float	Euro	Required	The deviation in the maintenance costs of yard trucks, based on the predictive maintenance tool.
Part type	String	-	Required	The part type used for maintenance
Type	String	-	Required	To distinguish historical logged data from trial data.
Start date	String	-	Required	Start date measurements of maintenance cost.
End date	String	-	Required	End date measurements of maintenance cost.

3.4.2 Hamburg LL data description

The Hamburg LL has specified a common data model that describes the data that will be handed over to the central tool.

Each KPI data set is assigned to a condition ID and location ID that describe the current traffic conditions and the location of the reported data set.

Each location ID represents one segment of the test field for autonomous and connected driving in Hamburg (TAVFTAVF) where the trials will take place. As the segmentation is highly dependent on the latest traffic situation (i.e. road works, blocked roads and other temporary restrictions) the segments have not been defined yet. So, the mapping of location ID and road segment will be given later.

The condition ID represents the traffic situation on Hamburg´s road network as delivered by the Hamburg authorities. The ID is related to one of 4 condition classes used to describe the current traffic situation while the traffic situation is determined by a speed index (GI). So, each data set will be assigned to a condition ID where:

- ID 1 is defined as class one, flowing traffic, GI > 0.7
- ID 2 is defined as class two, dense traffic, GI >= 0.4
- ID 3 is defined as class three, heavy traffic, GI >= 0.2
- ID 4 is defined as class four, congested traffic, GI < 0.2

Avg. truck speed single mode (H-KPI1)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Speed	Float	m/s	Required	Average Speed
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Avg. acceleration activities single mode (H-KPI2)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Acceleration	Float	m/s^2	Required	Average Acceleration
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Avg. stillstand time single mode (H-KPI3)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Stillstand time	Float	s	Required	Stillstand time spent in area
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Truck speed profile by platoon mode (H-KPI4)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Speed	Float	m/s	Required	Average Speed
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation

Acceleration profile by platoon mode (H-KPI5)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Acceleration	Float	m/s^2	Required	Average Acceleration
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int			External conditions e.g. influencing the traffic situation
LocationID	int		Required	Location where this speed has been recorded

Stillstand time profile by platoon mode (H-KPI6)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Stillstand time	Float	s	Required	Stillstand time spent in area
Total Time	Float	s		Total time spent in area
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Fuel consumption single mode (H-KPI7)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Fuel consumption	Float	l/100km	Required	Average fuel consumption
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

CO2 emissions single mode (H-KPI8)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
CO2 Emission	Float		Required	Average CO2 emission
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Fuel consumption platoon mode (H-KPI9)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Fuel consumption	Float		Required	Average fuel consumption
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

CO_2 emissions platoon mode (H-KPI10)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
CO_2 Emission	Float	g/km	Required	Average CO_2 emission
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int			External conditions e.g. influencing the traffic situation
LocationID	int		Required	Location where this speed has been recorded

Energy performance index value EPI (H-KPI11)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
EPI	Float	l/100km*t	Required	Energy performance index
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Acceleration performance index value API (H-KPI12)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
API	Float	kWh/100km*t	Required	Acceleration performance index
StartDate	Date	-	Required	Start date and time of this test drive
EndDate	Date	-	Required	End date and time of this test drive
IncludingGLOSA	Boolean	-		GLOSA in use at this test drive, YES or NO
VehicleType	String	-		Model and type of the test vehicle
Conditions	int	-		External conditions e.g. influencing the traffic situation
LocationID	int	-	Required	Location where this speed has been recorded

Available 5G bandwidth on urban roads (H-KPI13)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Timestamp	Date	-	Required	
Reference Signal Received Power	Float	dBm		Reference Signal Received Power (RSRP) in dBm
Reference Signal Received Quality	Float	dB		Reference Signal Received Quality (RSRQ) in dB
Upload Rate	Float	Mbps	Required	Upload Rate in Mbit/s
Download Rate	Float	Mbps	Required	Download Rate in Mbit/s
Location information	GPS			Location where this value has been tested

Positioning quality on urban road networks with 5G (H-KPI14)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Timestamp	Date	-	Required	
GPS accuracy	Float		Required	GPS accuracy without correction
Corrected accuracy	Float		Required	Accuracy with correction

Latency by 5G cellular communication in urban areas (H-KPI15)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Timestamp			Required	
Reference Signal Received Power		dBm		Reference Signal Received Power (RSRP) in dBm
Reference Signal Received Quality		dB		Reference Signal Received Quality (RSRQ) in dB
Upload Latency Edge		ms	Required	Upload Latency between device and Edge in ms
Download Latency Edge		ms	Required	Download Latency between Device and Edge in ms
Upload Latency Cloud		ms	Required	Upload Latency between device and Cloud in ms
Download Latency Cloud		ms	Required	Download Latency between Device and Cloud in ms
Location information				Location where this value has been tested

3.4.3 Koper LL data description

According to the storyboards, the Koper LL will hand over the following data to the central data collection tool.

3.4.3.1 Technical/5G KPI data

Components Onboarding and Configuration Time - 5G IoT backend (K-KPI1)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
OnboardingTime	Int	s	required	Elapsed time from the beginning of component configuration and onboarding process via the orchestrator until the components are ready to deploy
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	server host id where the components were deployed

Deployment Time - 5G IoT backend (K-KPI2)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
DeploymentTime	int	s	required	Elapsed time from the moment the deployment is started via the orchestrator until the system is ready to use
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	server host id where the components were deployed

Time to scale - 5G IoT backend (K-KPI3)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
TimeToScale	int	s	required	Elapsed time from the moment the scaling request is triggered until the component is scaled and ready to use
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	server host id where the components were deployed

Service Availability - 5G IoT backend (K-KPI4)				
Field	**Type**	**Unit**	**Required /Optional**	**Description**
Availability	float	%	required	Percentage of successful connection tests (RTT) and service tests (WEB) to the reference service endpoint over a period of time
Type	int	-	required	Service (1), Network (2)
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
AgentId	String	-	optional	Id of qMON agent doing service health checks
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	Server host id where the components were deployed

Components Onboarding and Configuration Time - 5G IoT agent (K-KPI5)				
Field	**Type**	**Unit**	**Required /Optional**	**Description**
Onboarding Time	Int	s	required	Elapsed time from the beginning of component configuration and onboarding process via the orchestrator until the components are ready to deploy
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	Server host id where the components were deployed

Deployment Time - 5G IoT agent (K-KPI6)				
Field	**Type**	**Unit**	**Required /Optional**	**Description**
Deployment Time	int	s	required	Elapsed time from the moment the deployment is started via the orchestrator until the system is ready to use
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	Server host id where the components were deployed

Components Onboarding and Configuration Time - 5G CN and 5G BBU (K-KPI7)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
OnboardingTime	int	s	required	Elapsed time from the beginning of component configuration and onboarding process via the orchestrator until the components are ready to deploy
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	Server host id where the components were deployed

Deployment Time - 5G CN and 5G BBU (K-KPI8)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
DeploymentTime	int	s	required	Elapsed time from the moment the deployment is started via the orchestrator until the system is ready to use
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	Server host id where the components were deployed

Time to scale - 5G CN and 5G BBU (K-KPI9)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
TimeToScale	int	s	required	Elapsed time from the moment the scaling request is triggered until the component is scaled and ready to use
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
Timestamp	Datetime /UTC	-	required	
HostID	String	-	optional	Server host id where the components were deployed

Service Availability - 5G CN and 5G BBU (K-KPI10)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Availability	float	%	required	Percentage of successful connection tests (RTT) and BW tests (Iperf) to the reference service endpoint over a period of time
Type	int	-	required	Service (1), Network (2)
ModuleId	Int	-	required	5G IoT backend (1), 5G IoT agent (2), 5G CN and 5G BBU (3)
AgentId	String	-	optional	Id of qMON agent doing service health checks
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	Server host id where the components were deployed

Slice Reconfiguration - 5G CN and 5G BBU (K-KPI11)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
SliceReconfigureTime	int	s	required	Elapsed time from the moment the slice reconfiguration is requested until the slice is reconfigured and ready to use
ModuleId	Int	-	required	5G CN and 5G BBU (3)
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken
HostID	String	-	optional	server host id where the slice was reconfigured

5G Network Area Traffic Capacity (K-KPI12)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
AreaTrafficCapacity	float	bps/m2	required	The total traffic throughput served per geographic area (in bps/m2)
LocationLatitude	float	degree	required	Geographical latitude coordinates
LocationLongitude	float	degree	required	Geographical longitude coordinates
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken

5G Network Connection Density (K-KPI13)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
ConnectionDensity	int	Per km2	required	Total number of connected and/or accessible devices per unit area (per km2)
LocationLatitude	float	Degree	required	Geographical latitude coordinates
LocationLongitude	float	Degree	required	Geographical longitude coordinates
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken

5G Network Coverage Area Probability (K-KPI14)

Field	Type	Unit	Required/Optional	Description
CoverageAreaProbability	float	%	required	Percentage (%) of the area under consideration, in which a service is provided by the mobile radio network to the end user in a quality that is sufficient for the intended application
LocationLatitude	float	Degree	required	Geographical latitude coordinates
LocationLongitude	float	Degree	required	Geographical longitude coordinates
Timestamp	Datetime/UTC	-	required	Time when the measurement was taken

5G Network Availability (K-KPI15)

Field	Type	Unit	Required/Optional	Description
Availability	float	%	required	Percentage of successful connection tests (RTT)/ service tests (WEB) to the reference service endpoint over a period of time
Type	int	-	required	Service (1), Network (2)
TestType	Int	-	required	Continuous (1), Drive (2)
AgentId	String	-	optional	id of qMON agent doing service health checks
Target	IP	-	optional	IP address of test end point (any public or private IPv4 address)
LocationLatitude	float	Degree	required	Geographical latitude coordinates
LocationLongitude	float	Degree	required	Geographical longitude coordinates
Timestamp	Datetime/UTC	-	required	Time when the measurement was taken

5G Network Bandwidth (K-KPI16)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Bandwidth	float	bps	required	Uplink and downlink bandwidth measured from the end user device (5G UE) on 5G RAN to the reference server located in 5G core
Type	String	-	required	Upload (UL), Download (DL)
TestType	Int	-	required	Continuous (1), Drive (2)
AgentId	String	-	optional	id of qMON agent doing bandwidth measurements
Target	IP	-	optional	IP address of test end point (any public or private IPv4 address)
LocationLatitude	float	Degree	required	Geographical latitude coordinates
LocationLongitude	float	Degree	required	Geographical longitude coordinates
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken

5G Network End-to-End Latency (K-KPI17)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Latency	int	ms	required	Measured round trip time (RTT) from the moment the ICMP Echo Request packet leaves the source host until the IP ICMP Echo Reply is received from the destination host
TestType	Int	-	required	Continuous (1), Drive (2)
AgentId	String	-	optional	id of qMON agent doing bandwidth measurements
Target	IP	-	optional	IP address of test end point (any public or private IPv4 address)
LocationLatitude	float	Degree	required	Geographical latitude coordinates
LocationLongitude	float	Degree	required	Geographical longitude coordinates
Timestamp	Datetime /UTC	-	required	Time when the measurement was taken

5G Network Reliability (K-KPI18)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Reliability	int	%	required	the percentage (%) of the amount of sent network layer packets successfully delivered to a given system node (incl. the 5G UE) within the time constraint required by the targeted service, divided by the total number of sent network layer packets
TestType	Int	-	required	Continuous (1), Drive (2)
AgentId	String	-	optional	id of qMON agent doing bandwidth measurements
Target	IP	-	optional	IP address of test end point (any public or private IPv4 address)
LocationLatitude	float	Degree	required	Geographical latitude coordinates
LocationLongitude	float	Degree	required	Geographical longitude coordinates
Timestamp	Datetime/ UTC	-	required	Time when the measurement was taken

3.4.3.2 Technical/Image processing KPI data

Model accuracy (K-KPI19) – (UC5)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Confusion matrix	Int	-	Required	The number of correctly detected damages (true positives and false negatives) and the badly detected damages (true negatives and false positives) by the UC5 developed ML model (evaluation data: set of manually annotated real images) compared with the number of manually annotated damages.

Model inference time (K-KPI20) – (UC5)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Manual time	Float	ms	Required	The time required to visually detect the presence/absence of damages and the detection of IMDG label (baseline data: values based on the operator observation).
Inference time	Float	ms	Required	The time required to process the input of video stream(s) and infer the presence/absence of damages and the IMDG labels based on UC5 developed ML model (trial data: synthetic images and manually annotated real images)
Standard deviation	Float	ms	Required	The standard deviation of the average values measured at the trials.

Model accuracy (K-KPI19) – (UC5)				
Field	**Type**	**Unit**	**Required/ Optional**	**Description**
Confusion matrix	Int	-	Required	The number of correctly detected people (true positives and false negatives) and the badly detected people (true negatives and false positives) by the UC6 developed ML model compared with the ground truth (set of manually annotated real images or set of open available videos in the literature review).

Model inference time (K-KPI22) – (UC6)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Theoretical time	Float	ms	Required	The time required to process the input of video stream(s) and infer the presence/absence of people (baseline data: statistical values based on literature review).
Inference time	Float	ms	Required	The time required to process the input of video stream(s) and infer the presence/absence of people at the selected area based on UC6 developed ML model.
Standard deviation	Float	ms	Required	The standard deviation of the average values measured at the trials.

Model Accuracy (K-KPI23) – (UC6)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Confusion matrix	Int	-	Required	The number of correctly detected vehicles (true positives and false negatives) and the badly detected vehicles (true negatives and false positives) by the UC6 developed ML model compared with the ground truth (set of manually annotated real images or set of open available videos in the literature review).

Model inference time (K-KPI24) – (UC6)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Theoretical time	Float	ms	Required	The time required to process the input of video stream(s) and infer the presence/absence of vehicle (baseline data: statistical values based on literature review).
Inference time	Float	ms	Required	The time required to process the input of video stream(s) and infer the presence/absence of vehicle at the selected area based on UC6 developed ML model.
Standard deviation	Float	ms	Required	The standard deviation of the average values measured at the trials.

3.4.3.3 Logistics KPI data

Time with engine stopped (K-KPI25)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
TimeStopped	int	s	Required	Time spent with engine stopped, over the given period
StartDate	Date	-	Required	
EndDate	Date	-	Required	
VehicleType	String	-	Required	Vehicle make and model on which IoT device is installed on. This allow making accurate comparisons between similar vehicles

Truck speed (K-KPI26)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Speed	Float	km/h	Required	Average Speed
StartDate	Date	-	Required	
EndDate	Date	-	Required	
VehicleType	String	-	Required	Vehicle make and model on which IoT device is installed on. This allow making accurate comparisons between similar vehicles
Conditions	int	-		External conditions e.g. influencing the traffic situation (i.e. weather conditions, night vs. Day etc.)
LocationID	int	-	Required	Location where this speed has been recorded

Truck Acceleration (K-KPI27)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
Acceleration	Float	m/s^2	Required	Average acceleration, in m/s^2, over the given period
StartDate	Date	-	Required	
EndDate	Date	-	Required	
VehicleType	String	-	Required	Vehicle make and model on which IoT device is installed on. This allow making accurate comparisons between similar vehicles
Conditions	int	-		External conditions e.g. influencing the traffic situation (i.e. weather conditions, night vs. Day etc.)
LocationID	int	-	Required	Location where this speed has been recorded

Truck Stand Still Time (K-KPI28)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
TimeInIdle	int	s	Required	Time spent in idle, over the given period
StartDate	Date	-	Required	
EndDate	Date	-	Required	
VehicleType	String	-	Required	Vehicle make and model on which IoT device is installed on. This allow making accurate comparisons between similar vehicles

Fuel Consumption (K-KPI29)				
Field	**Type**	**Unit**	**Required/Optional**	**Description**
FuelConsumption	Float	l/h	Required	Average fuel consumption, in the specified period of time
StartDate	Date	-	Required	
EndDate	Date	-	Required	
VehicleType	String	-	Required	Vehicle make and model on which IoT device is installed on. This allow making accurate comparisons between similar vehicles
Conditions	int	-		External conditions e.g. influencing the traffic situation (i.e. weather conditions, night vs. Day etc.)
LocationID	int	-	Required	Location where this speed has been recorded

4 LIVING LABS DATA COLLECTION TOOLS

4.1 Athens Living Lab

All data will be collected by the Athens Living Lab partners and handed over to the data collection tool based on the Fluentd architecture (protocols, interfaces, etc.), according to the respective data models (and data formats) that will be selected. For KPI sensitive data (due to business restrictions and privacy/security issues) the deviation of baseline (i.e., prior to 5G-LOGINNOV) and trial data will be handed over to the data collection tool, in order to evaluate the use case impact on the LL targeted operations. For use cases that correspond to 5G connected yard trucks (UC3 and UC7), the data sources include: telemetry data (CAN-Bus, localization, container presence and other custom sensor data) from each yard truck, transmitted to the PCT management platform via the integrated (on-truck) 5G telematics device; the traffic monitoring system (TMS) which exploits the aggregated telemetry data from the fleet of 5G trucks to provide logged/historical, operational and use case data of daily port operations for business and analytics; as well as the Enterprise Asset Management System (EAMS) at PCT datacentre, which includes information regarding the description of vehicles operational status (historical and recent data), e.g., hours of operation, breakdown events and duration, parts of the truck that were affected by a malfunction and the spare parts used for the repair etc. Both systems (TMS, EAMS) are coupled with their dedicated database that can be queried for data retrieval. For UC2 the operational data will be collected and processed at the Vodafone Innovus platform which will also be linked to the Fluentd agent.

For use cases based on NFV-MANO orchestration of video analytics tasks to 5G-IoT devices (UC4 and UC5), metrics regarding the performance of the computer visions tasks (e.g., machine learning model inference time and accuracy) will be stored at the edge-device and pulled to the MANO orchestrator (based on open-source MANO - OSM) on specified intervals. Similarly, MANO related KPIs will be monitored and logged at the OSM controller, and later linked to the Fluentd client. Tools such as iperf and ping will be used to measure the 5G-IoT device experienced data rate (uplink/downlink) for massive uplink video data transmissions as well as for latency measurement, logged by either the edge device or the OSM controller. For the 5G NSA network established by the Greek pilot’s MNO (Vodafone), dedicated tools will be exploited covering the entire range of the Athens LL 5G KPIs (latency, bandwidth, etc.) to ensure the efficient operation of the cellular network at PCT premises.

Overall, all use case (operational and technical) data will be aggregated at a local PCT (and/or ICCS) server, linked to the external Fluentd client for sending all relevant data (according to the agreed data models and formats, e.g., JSON) to the central data collection tool for the evaluation of the Greek pilot. Figure 6 summarizes the flow of data from end devices to the Fluend client.

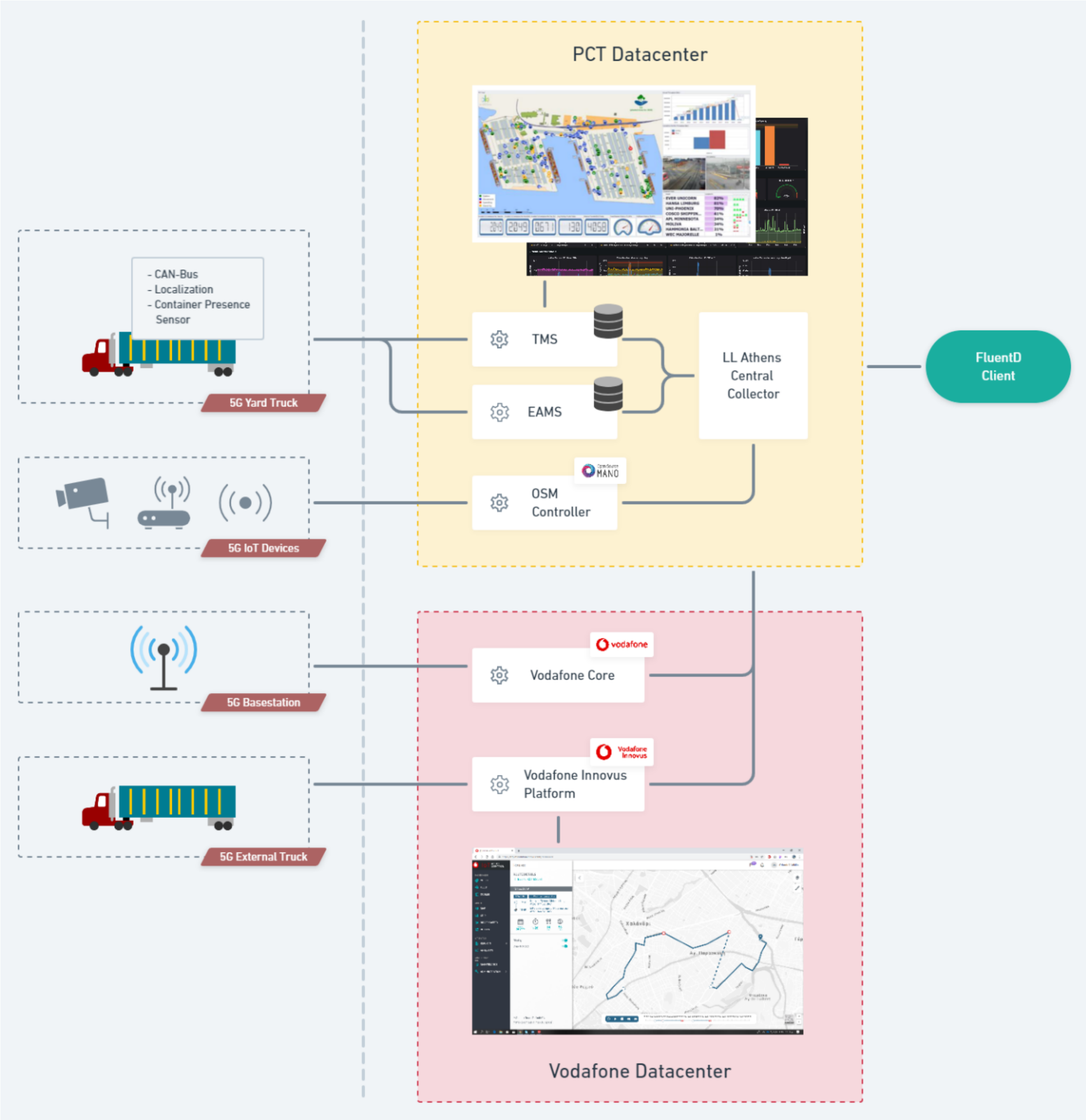

Figure 6: Data collection architecture – Athens Living Lab

4.2 Hamburg Living Lab

The Hamburg Living Lab will mostly provide processed data to the central data collection tool. As long as applicable, data available from all Living Lab partners (Continental, tec4u, SWARCO and T-Systems) will be collected and aggregated in a Living Lab data collection instance whose main purpose lies in acting as the main interface to transfer data to the central data collection instance via Fluentd. As the data foreseen for the evaluation tool is the result of complex algorithms it is not applicable for detailed processing by the evaluation tool and is only intended for limited aggregation and visualisation.

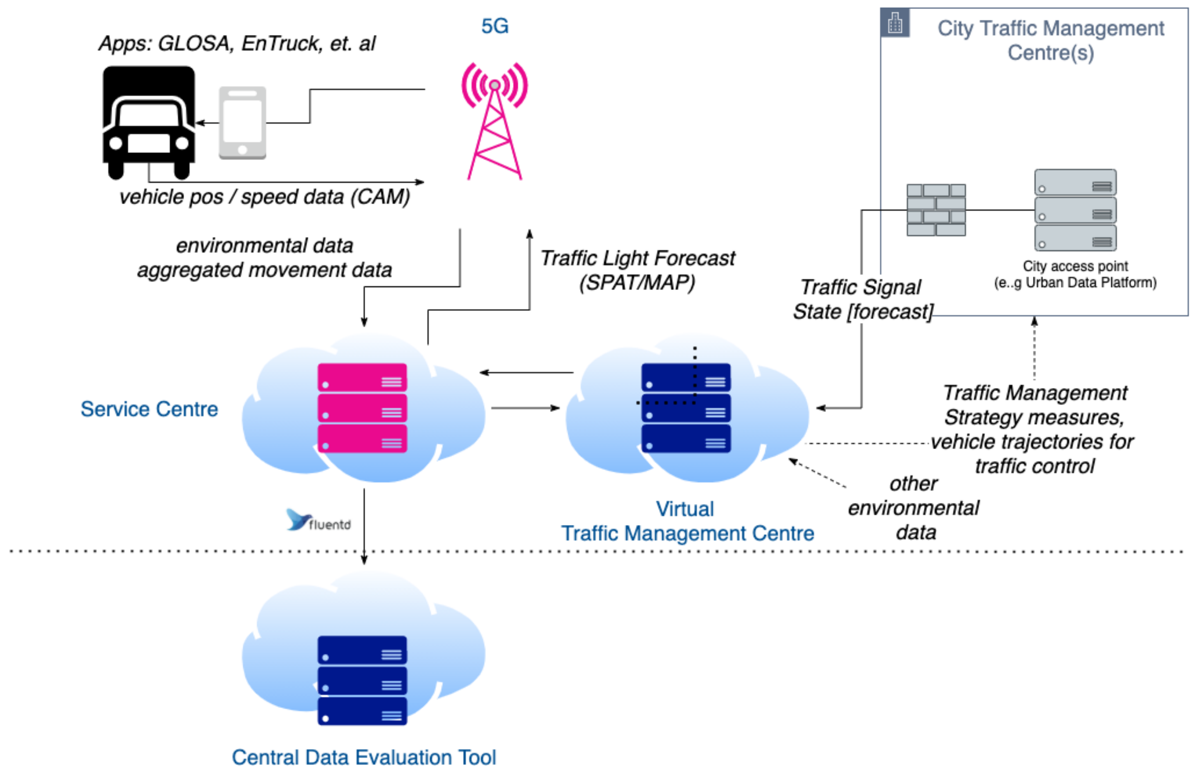

Figure 7: Data collection architecture – Hamburg Living Lab

4.3 Koper Living Lab

Koper Living Lab facility will provide centralized monitoring point based on Fluentd which will provide multiple inputs on one side and act as a single output to 5G-LOGINNOV's central data collection tool.

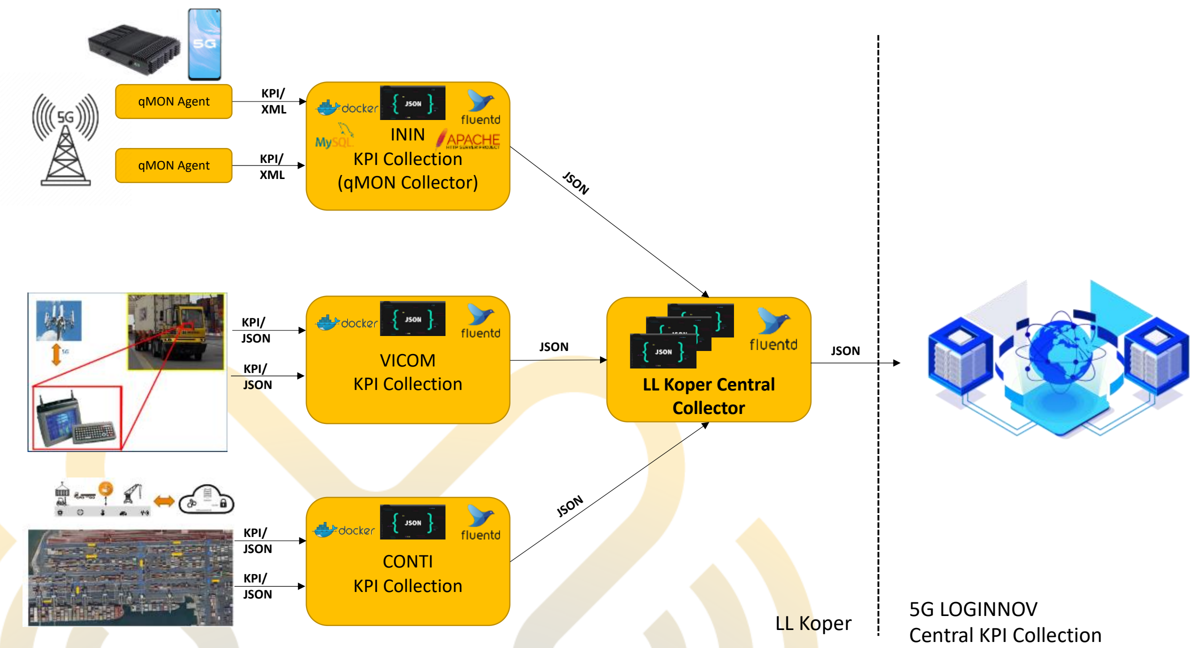

Figure 8: Data collection architecture – Koper Living Lab

Data sources for multiple inputs to the Koper LL central collector will be realized with Fluentd-based KPI collectors that will be provisioned per use case partner (i.e. ININ, VICOM, CONTI). In the case of

ININ KPI collector, the existing qMON Collector (responsible for gathering and parsing KPIs from qMON Agents) entity will be extended to support the extraction of relevant 5G-LOGINNOV KPIs from the native qMON KPI datasets to Fluentd supported format. For the other two partners, dedicated Docker containers will be provisioned to allow the Fluentd agents processing the files from the partner's use case scenarios. This way, the only requirements from use case partners are to provide JSON-based files containing relevant 5G-LOGINNOV KPIs and be able to put those files in selected folders (as shown in the figure).

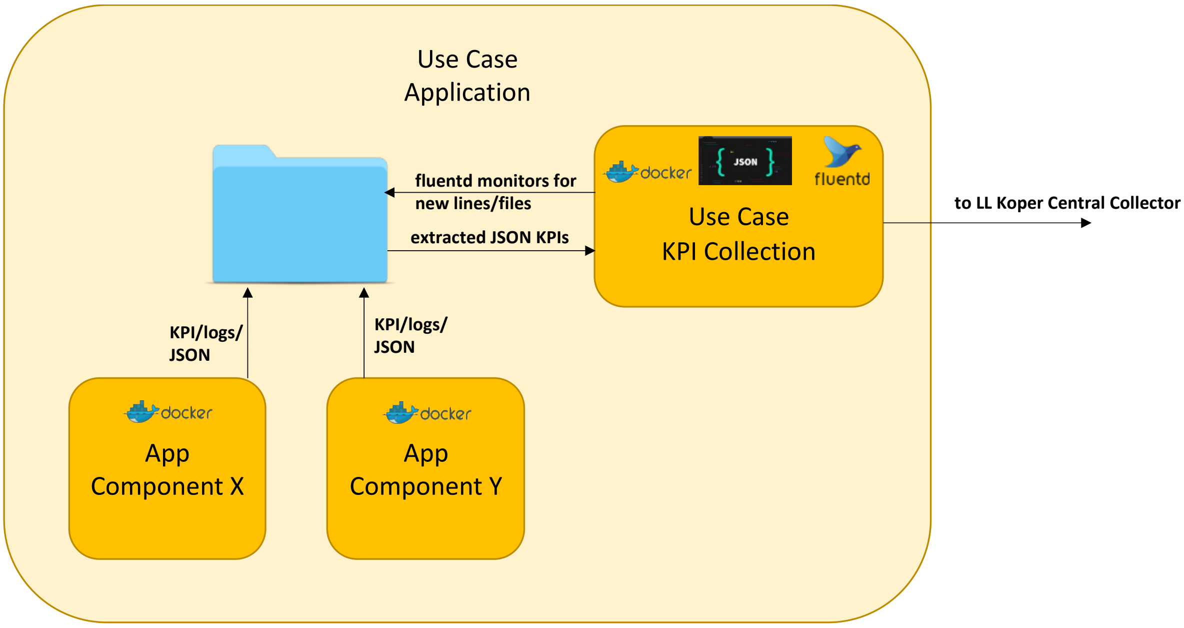

Figure 9: KPI collection using Docker – Koper Living Lab

5 CENTRAL DATA COLLECTION TOOL

This chapter describes the implementation of the central data collection tool, whose role is to ingest and store the evaluation data from the Living Labs making them available to the evaluation team.

5.1 Components

The central data collection tool relies on a set of tools from open-source projects as illustrated on Figure 10. The design of this architecture of the central data collection tool was based on the requirements specified earlier (2.1), which drove the choice of the components and modules with their roles.

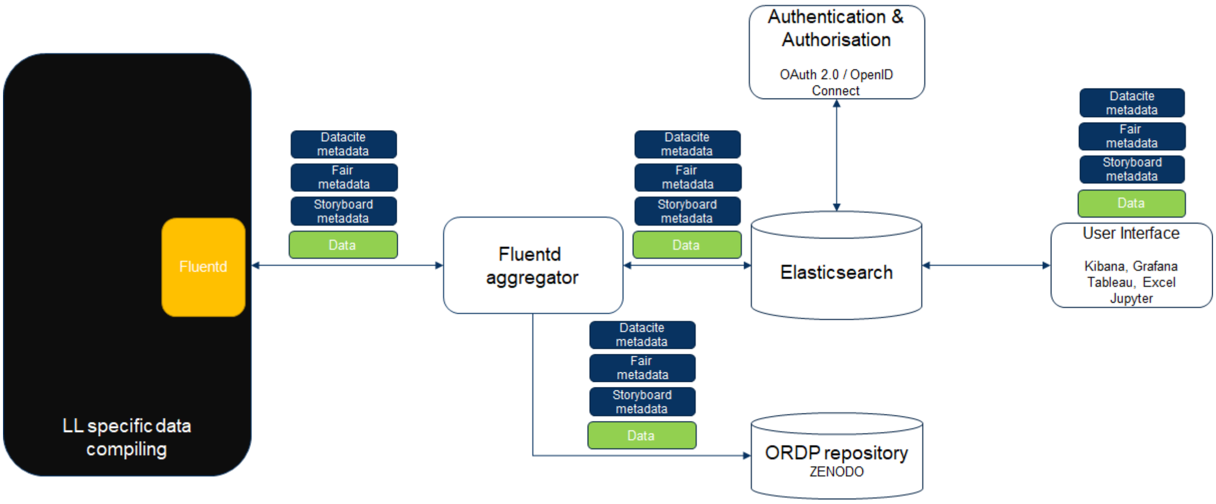

Figure 10: Central data collection tool components

The open-source projects used for the implementation of the central data collection tools are presented in the Table 7.

Name	Type (or Requirement ID)	Description
FluentD	Communication protocol from Local server to Central Server (LR-04)	Open-source project used in 5G-LOGINNOV as data ingestion engine and interface with the Living Labs.
TLS	Security protocol (CR-10)	Let's Encrypt TLS certificates are used in 5G-LOGINNOV to provide transport layer security.
Elastic Stack	Central Server Technology (CR-11, ER-05, CTR-XX)	Kibana and Elasticsearch are used for respectively providing a user interface and a data base with indexing and search.
Keycloak	Authentication and Authorization (CTR-02)	Open-source implementation of OAuth 2.0 and OpenID Connect among others authentication protocols. Keycloak is used in 5G-LOGINNOV to provide authentication and authorisation in servers such as Elasticsearch and Kibana.

Table 7: Open-source projects used to implement the central data collection tool

The data collections tools are deployed on Microsoft Azure.

5.2 Implementation

5.2.1 Data ingestion

5.2.1.1 Introduction to Fluentd

Fluentd[3] is an open-source project which is widely used as data collector. It aims to:

- Unify the logging with a common JSON data structure
- Offer a pluggable architecture based on data sources, data filters and data outputs
- Use as minimum resources as possible
- Offer a built-in reliability

The log unification of Fluentd is based on the usage of JSON to harmonise different types of output as illustrated on Figure 11.

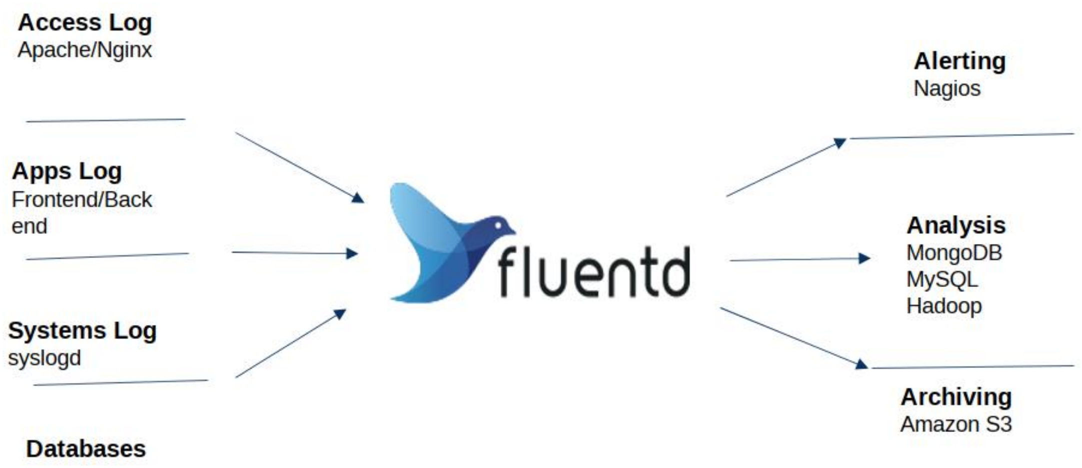

Figure 11: Common logging provided by Fluentd

Fluentd provides a set of pluggable modules which ensure that different data sources can be used, filtered and sent to different outputs. Figure 12 shows examples of the pluggable modules in the form of filters that are applied to the incoming or outgoing data.

[3] https://www.fluentd.org/

Certified	Download	Name	Author	About	Version
	26440076	rewrite-tag-filter	Kentaro Yoshida	Fluentd Output filter plugin. It has designed to rewrite tag like mod_rewrite. Re-emmit a record with rewrited tag when a value matches/unmatches with the regular expression. Also you can change a tag from apache log by domain, status-code(ex. 500 error), user-agent, request-uri, regex-backreference and so on with regular expression.	2.4.0
	12834112	record-modifier	Masahiro Nakagawa	Filter plugin for modifying event record	2.1.0
	5787169	concat	Kenji Okimoto	Fluentd Filter plugin to concat multiple event messages	2.5.0
	4424063	kubernetes_metadata_filter	Jimmi Dyson	Filter plugin to add Kubernetes metadata	2.9.0
	4070153	throttle	François-Xavier Bourlet	Fluentd filter for throttling logs based on a configurable key.	0.0.5
	3066345	geoip-filter	Yuri Umezaki	Fluentd filter plugin to add geoip	1.0.0
	2906601	parser	TAGOMORI Satoshi	fluentd plugin to parse single field, or to combine log structure into single field	0.6.1
	2146916	script	SNakano	Fluentd filter plugin to external ruby script	0.1.1
	1937145	anonymizer	Kentaro Yoshida	Fluentd filter output plugin to anonymize records with HMAC of MD5/SHA1/SHA256/SHA384/SHA512 algorithms. This data masking plugin protects privacy data such as UserID, Email, Phone number, IPv4/IPv6 address and so on.	1.0.0
	1298705	geoip	Kentaro Yoshida	Fluentd Filter plugin to add information about geographical location of IP addresses with Maxmind GeoIP databases.	1.3.2

Figure 12: Fluend extensions modules as Filters to manipulate logs

5.2.1.2 Installation and Configuration

The installation, configuration and orchestration of Fluentd is done through Ansible[4] modules and Ansible playbooks. The current version of the data collection tool is using Fluentd version v1.13.3 from the Treasure agent (td-agent) distribution package.

[4] https://www.ansible.com/

```yaml
---
# tasks file for fluentd
#
- name: Add td-agent GPG keys
  apt_key:
   url: "{{ treasures_data_gpg_keys_url }}"
   state: present

- name: Add Treasure Data repository
  apt_repository:
   repo: "{{ treasures_data_dpkg_repository }}"
   state: present
   filename: treasure-data

- name: Install td-agent and "acl"
  apt:
   name: ["td-agent", "acl"]
   state: present
   update_cache: yes

- name: Increase the Maximum Number of File Descriptors
  copy:
    src: limits.conf
    dest: /etc/security/limits.conf

- name: Optimize the Network Kernel Parameters
  copy:
    src: 10-fluentd.conf
    dest: /etc/sysctl.conf.d/

- name: Copy Fluentd configuration  files
  copy:
    src: files/td-agent
    dest: /etc/
```

Figure 13: Installation and configuration of Fluentd using Ansible

5.2.1.3 Interface towards the Living Labs

The interfacing of the central data collection tool with the Living Labs relies on Fluentd's Forward protocol[5] which uses MessagePack[6] data format. It allows the instances of Fluentd to communicate between themselves. An example of MessagePack data is illustrated in Figure 14.

[5] https://github.com/fluent/fluentd/wiki/Forward-Protocol-Specification-v1
[6] https://msgpack.org/

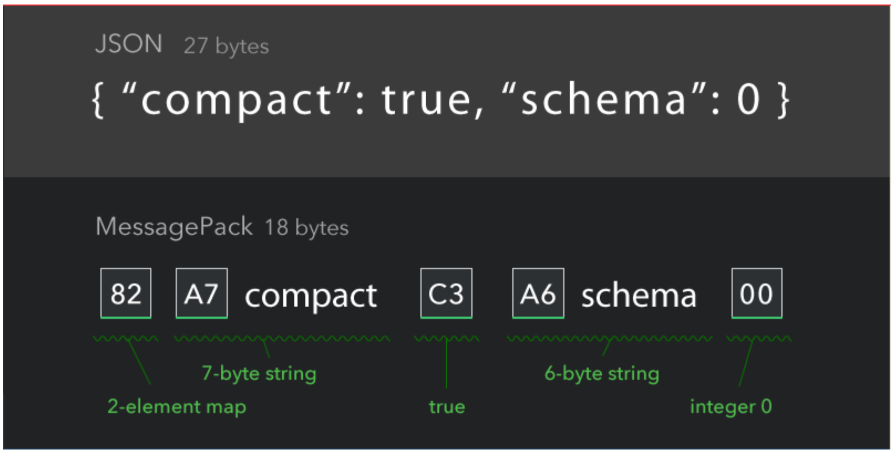

Figure 14: MessagePack data format

An excerpt of Fluentd configuration for the central tool – Living Labs interface is given in Figure 15. It highlights the use of TLS for secure communications and the need for the Living Labs to authenticate.

```
<source>
  @type forward
  bind 0.0.0.0
  send_keepalive_packet true

  <transport tls>
        cert_path /etc/letsencrypt/live/dataingestion5gl.francecentral.cloudapp.azure.com/fullchain.pem
        private_key_path /etc/letsencrypt/live/dataingestion5gl.francecentral.cloudapp.azure.com/privkey.pem
        private_key_passphrase ""
        client_cert_auth false
  </transport>

  <security>
    self_hostname dataingestion5gl.francecentral.cloudapp.azure.com
    shared_key ""
  </security>

</source>
```

Figure 15: Security configuration of Fluentd to use TLS and to authenticate LLs

Fluentd listens on port 24224. Upon receiving the data, it applies certain filters as described in Figure 2.

5.2.1.4 Interface with the database

Once the data are validated, Fluentd forwards them to Elasticsearch using an output called 'Elasticsearch'[7], which uses the Elasticsearch REST APIs (described in 5.2.2.4). Fluentd uses a client developed in Ruby to access the APIs. The corresponding Fluentd configuration is shown in Figure 16.

[7] https://github.com/uken/fluent-plugin-elasticsearch

```
# Forward data from Athens to Elasticsearch
<match ll-athens>
  @type elasticsearch
  host database5gl.francecentral.cloudapp.azure.com
  scheme https
  user elastic
  password ""
  port 9200
  index_name ll-athens
  type_name  ll-athens
</match>

# Forward data from Hambourg to Elasticsearch
<match ll-hambourg>
  @type elasticsearch
  host database5gl.francecentral.cloudapp.azure.com
  scheme https
  user elastic
  password ""
  port 9200
  index_name ll-hambourg
  type_name  ll-hambourg
  template_name ll-hamboug

</match>

# Forward data from Koper to Elasticsearch
<match ll-koper>
  @type elasticsearch
  host database5gl.francecentral.cloudapp.azure.com
  scheme https
  user elastic
  password ""
  port 9200
  index_name ll-koper
  type_name  ll-koper
</match>
```

Figure 16: Fluentd configuration to send the evaluation data to Elasticsearch

5.2.1.5 Interface with ORDP repositories

Evaluation data will be sent to an ORDP repository (ZENODO is recommended in the D6.4 [2]) using Fluentd. This is done by configuring an output in Fluentd configuration using REST APIs or the protocol supported by the chosen repository. Figure 17 shows an example of Zenodo API that can be used by Fluentd.

```
{
    "conceptrecid": "542200",
    "created": "2020-05-19T11:58:41.606998+00:00",
    "files": [],
    "id": 542201,
    "links": {
        "bucket": "https://zenodo.org/api/files/568377dd-daf8-4235-85e1-a56011ad454b",
        "discard": "https://zenodo.org/api/deposit/depositions/542201/actions/discard",
        "edit": "https://zenodo.org/api/deposit/depositions/542201/actions/edit",
        "files": "https://zenodo.org/api/deposit/depositions/542201/files",
        "html": "https://zenodo.org/deposit/542201",
        "latest_draft": "https://zenodo.org/api/deposit/depositions/542201",
        "latest_draft_html": "https://zenodo.org/deposit/542201",
        "publish": "https://zenodo.org/api/deposit/depositions/542201/actions/publish",
        "self": "https://zenodo.org/api/deposit/depositions/542201"
    },
    "metadata": {
        "prereserve_doi": {
            "doi": "10.5072/zenodo.542201",
            "recid": 542201
        }
    },
    "modified": "2020-05-19T11:58:41.607012+00:00",
    "owner": 12345,
    "record_id": 542201,
    "state": "unsubmitted",
    "submitted": false,
    "title": ""
}
```


```
$ curl https://zenodo.org/api/deposit/depositions/222761?access_token=$ACCESS_TOKEN
{ ...
  "links": {
    "bucket": "https://zenodo.org/api/files/568377dd-daf8-4235-85e1-a56011ad454b",
    ...,
  },
... }
```

Figure 17: ZENODO API that will be used by Fluentd to push ORDP data

5.2.2 Database

5.2.2.1 Introduction to Elasticsearch

Elasticsearch[8] is the distributed search and analytics engine at the heart of the Elastic Stack[9] which provides near real-time search and analytics for all types of data. In 5G-LOGINNOV, Elasticsearch is used primarily as a database but other features such as searching are also provided.

5.2.2.2 Installation and Configuration

The installation and configuration of Elasticsearch uses Ansible as illustrated on Figure 18.

[8] https://www.elastic.co/
[9] https://www.elastic.co/elastic-stack/

```yaml
---
# tasks file for elastic
#
- name: Install and Elastic OpenJDK-14
  apt:
   name: ["openjdk-14-jre", "elasticsearch"]
   state: present

- name: Copy Elasticsearch configuration
  copy:
   src: files/elasticsearch
   dest: /etc/
   owner: elasticsearch
   group: elasticsearch

- name: Copy sysctl.conf file
  copy:
   src: files/sysctl.conf
   dest: /etc/sysctl.conf

- name: Update rc and start Elasticsearch
  service:
   name: elasticsearch
   state: started
   enabled: yes

```

Figure 18: Installation and configuration of Elasticsearch using Ansible

5.2.2.3 Design of the 5G-LOGINNOV's LL indices

Based on sections 3.3 and 3.4, respectively related to the metadata and the data from the Living Labs, the indices for each Living Lab have been designed using JSON and the data types provided by Elasticsearch. The main indices are:

- Athens index
- Hamburg index
- Koper index
- Evaluation data index

The detailed descriptions of the indices will be provided in the future deliverable D2.3, as they rely on some input from the Living Labs development that is still ongoing, as well as the evaluation tasks that are yet to start.

5.2.2.4 REST API

Elasticsearch provides a REST API that allows the management of the indices and access to the data stored in these indices. It will be used by the data ingestion components and the user interface described in 5.2.3. An example of API usage is shown in Figure 19.

The choice of the Elasticsearch REST API corresponds to the implementation of the ER-05 requirements related to the use of SQL-like requests which are convenient for the evaluation team.

```
PUT _component_template/ll-hambourg-test
{
  "template": {
  "mappings": {
      "dynamic": false,
      "properties": {
        "Time[s]": {
          "type": "date"
        },
        "AccECE[%]": {
          "type": "double"
        },
        "AccWork[J]": {
          "type": "double"
        },
        "Acceleration[m/s^2]": {
          "type": "double"
        },
        "AeroECE[%]": {
          "type": "double"
        },
        "AeroWork[J]": {
          "type": "double"
        },
        "Altitude[m]": {
          "type": "double"
        },
        "CO2[kg]": {
          "type": "double"
        },
        "Flag": {
          "type": "double"
        },
        "Fuel[l]": {
          "type": "double"
        },
        "GradeWork[J]": {
          "type": "double"
        },
        "location": {
          "type": "geo_point"
        },
        "RollWork[J]": {
          "type": "double"
        },
        "STSECE[%]": {
          "type": "double"
        },
        "Speed[m/s]": {
          "type": "double"
        },
        "StandStillTime[s]": {
          "type": "double"
        },
        "StandStillWork[J]": {
          "type": "double"
        },
        "TotalWork[J]": {
          "type": "double"
        },
        "WorkECE[%]": {
          "type": "double"
        }
      }
    }
  },
    "version": 1
}
}
```

Figure 19: Elasticsearch API as illustrated in Kibana

5.2.3 The user interface

A default user interface based on Kibana is available. However other solutions are supported such as Microsoft Excel[10], Jupyter[11] or Grafana[12].

5.2.3.1 Introduction to Kibana

Kibana[13] is an open-source project that is part of the Elastic stack. It uses the Elasticsearch API to manage Elasticsearch clusters and indices. Furthermore, it serves as a data visualisation tool as illustrated in Figure 20.

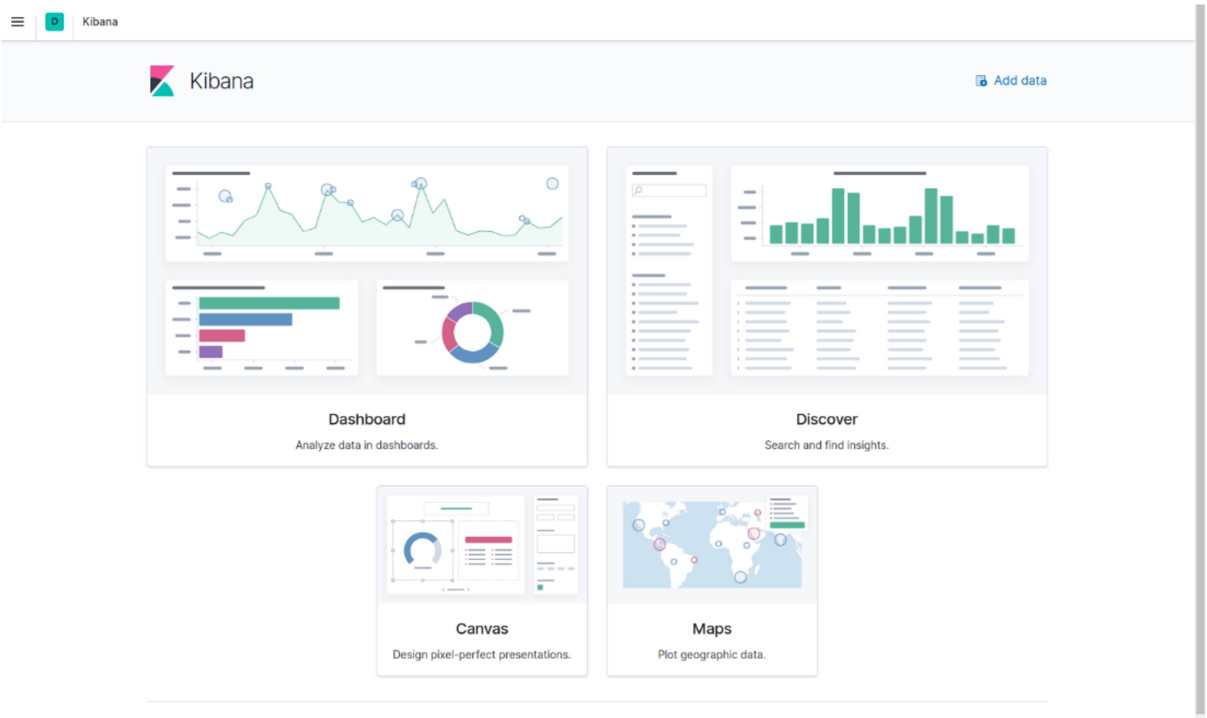

Figure 20: Kibana's UI for visualisation and management of dashboards and data

[10] https://www.microsoft.com/en/microsoft-365/excel
[11] https://jupyter.org/
[12] https://grafana.com/
[13] https://www.elastic.co/kibana/

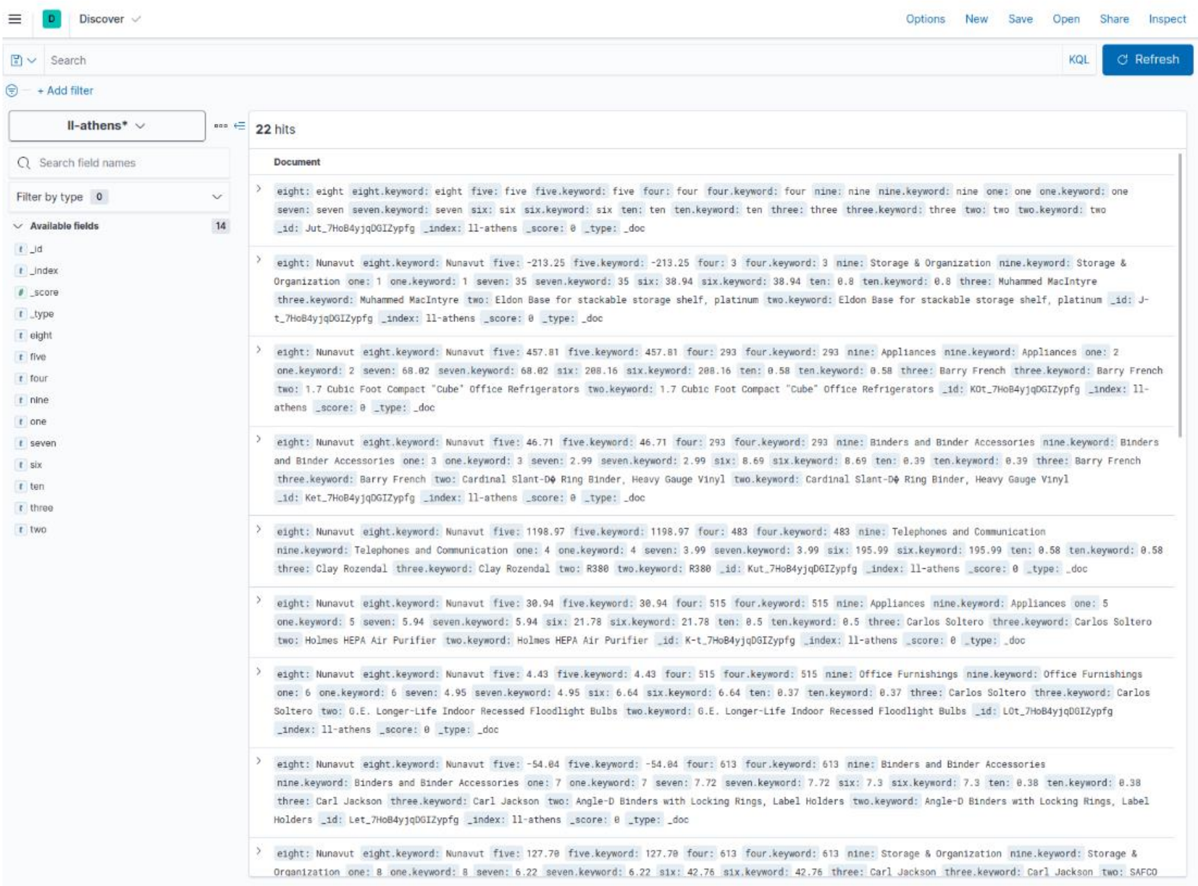

Figure 21: Sample data from Athens LL displayed in Kibana

5.2.3.2 Installation and configuration

The installation and configuration of Kibana are done using Ansible as illustrated for Fluentd and Elasticsearch in Figure 13 and Figure 18.

5.2.3.3 Interface with the evaluation team

The interfacing with the evaluation team can be done through the REST API described in section 5.2.2.4 or using the Kibana user interface. Additionally, other tools can be chosen as user interfaces to the database such as Excel, Jupyter, curl[14] etc.

Elasticsearch also supports the SQL language[15] usable through Kibana satisfying the ER-05 requirement.

5.2.4 Authentication and authorisation

The authentication and authorisation aspects of the central data collection tool rely on an implementation of OAuth 2.0 and OpenID Connect named Keycloak[16].

[14] https://curl.se/
[15] https://www.elastic.co/guide/en/elasticsearch/reference/current/xpack-sql.html
[16] https://www.keycloak.org/

5.2.4.1 Introduction to OAuth 2.0 and OpenID Connect

OAuth 2.0 [8] is an authorisation framework that allows a client application to access protected resources using access tokens provided by an authentication server. The protocol flow contextualised to 5G-LOGINNOV is illustrated in Figure 22.

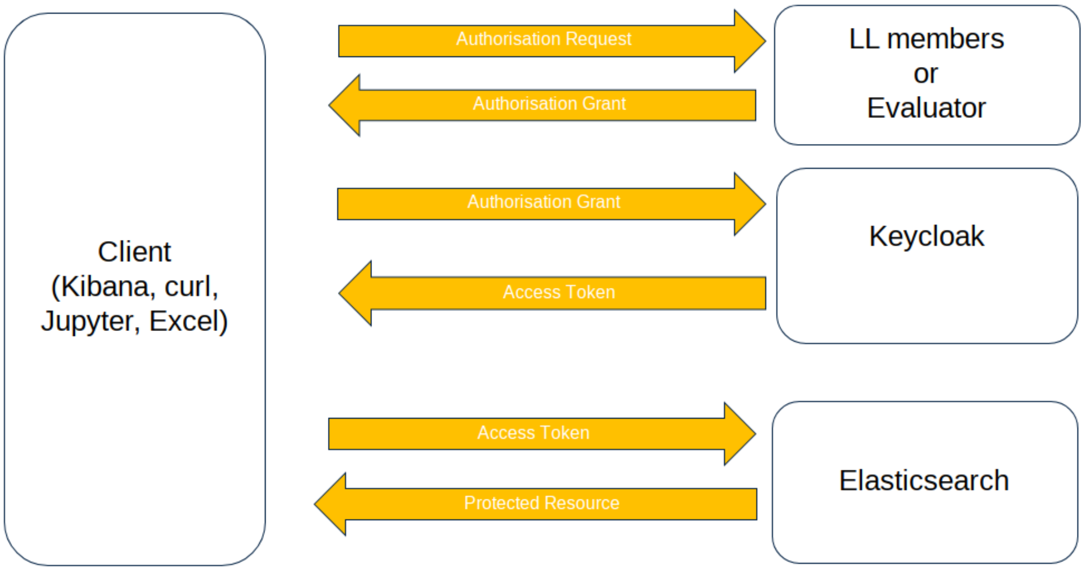

Figure 22: Authentication and authorisation process using OAuth 2.0

5.2.4.2 Installation and Configuration

The installation and of Keycloak is done using Ansible as illustrated for Fluentd and Elasticsearch in Figure 13 and Figure 18.

5.3 Deployment

The central data collection tool components are hosted on Microsoft Azure. The virtual machines configurations are described in Table 8.

Component	Specifications	Description
Data ingestion	vCPU: 2 RAM: 4Go Network: unlimited VM size: Standard_B2s	An Ubuntu VM used to host Fluentd and the necessary software dependencies such as Chrony.
Database	vCPU: 2 RAM: 8Go Network: unlimited VM size: Standard_B2ms	An Ubuntu VM used to host Elasticsearch and the necessary software dependencies such as JDK.
User interface	vCPU: 2 RAM: 4Go Network: unlimited VM size: Standard_B2s	An Ubuntu VM used to host Kibana and the necessary software dependencies such as JDK.

Authentication and authorisation	vCPU: 2 RAM: 4Go Network: unlimited VM size: Standard_B2s	An Ubuntu VM used to host Keycloak and the necessary software dependencies such as JDK.

Table 8: Central data collection tool components deployment

The deployment of the tools has been done using Ansible playbook as illustrated on Figure 23.

```yaml
---
- hosts: all
  gather_facts: false

- hosts: elastic
  roles:
    - common
    - certbot
    - elastic.co
    - elastic

  tasks:
    - name: Run Elasticsearch
      become: yes
      become_user: root
      service:
       name:  elasticsearch
       state: restarted

- hosts: fluentd
  roles:
    - common
    - certbot
    - fluentd

  tasks:

    - name: Start fluentd
      become: yes
      become_user: root
      service:
       name: td-agent
       state: restarted

- hosts: kibana
  roles:
    - common
    - certbot
    - elastic.co
    - kibana

  tasks:
    - name: Run Kibana
      become: yes
      become_user: root
      service:
       name:  kibana
       state: restarted
```

Figure 23: Deployment and orchestration of the central data collections tools on MS Azure using Ansible

6 CONCLUSION

This deliverable describes the tools that will be implemented and deployed for 5G-LOGINNOV evaluation data collection. The design of the data collection architecture that is specified in this document takes into consideration the requirements regarding the general data management of the project as well as evaluation and Living Labs specific requirements. Thanks to the architecture and the data management approach, including the description of the evaluation data to be transferred to the central data collection tool, the implementation and deployment choices for the tools have been made and are described in this document.

This document can be used by the Living Labs to get technical information regarding the central data collection tool such as the interfacing aspects. Additionally, this deliverable will support the evaluation tasks by providing metadata and evaluation data description to the evaluation team.

As the Living Labs development and deployment tasks are still ongoing and the evaluation tasks have not yet started by the writing of this deliverable, additional input is expected in the future that may require adaptation and updates on the tools' features. These updates will be described in D2.3 – "Development and deployment final report" which will be released in M20 (April 2022) of the project.

7 REFERENCES

[1] FOT-Net, “Field Operational Tests wiki,” [Online]. Available: https://wiki.fot-net.eu/. [Accessed 2021].

[2] 5G-LOGINNOV, “D6.4 Data Management Plan,” 2021.

[3] 5G-LOGINNOV, “D1.4 Initial specification of evaluation and KPIs,” 2021.

[4] 5G-LOGINNOV, “D1.5 Data and cyber-protection policies,” 2021.

[5] 5G-LOGINNOV, “D3.1 Trial methodology, planning and coordination,” 2021.

[6] M. Wilkinson, M. Dumontier and I. Aalbersberg, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci Data,* vol. 3, 2016.

[7] DataCite Metadata Working Group, *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs,* 2021.

[8] D. Hardt, “The OAuth 2.0 Authorization Framework,” 2012.